(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 November 2002 (21.11.2002)

PCT

(10) International Publication Number
WO 02/092801 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/EP02/05363

(22) International Filing Date: 15 May 2002 (15.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/291,149 16 May 2001 (16.05.2001) US

(71) Applicant *(for all designated States except US)*: **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MOLNAR, Istvan** [HU/US]; Syngenta Biotechnology, Inc., 3054 Cornwallis Road, Research Triangle Park, NC 27709 (US). **LIGON, James, Madison** [US/US]; Syngenta Biotechnology, Inc., 3054 Cornwallis Road, Research Triangle Park, NC 27709 (US). **ZIRKLE, Ross, Eric** [US/US]; Syngenta Biotechnology, Inc., 3054 Cornwallis Road, Research Triangle Park, NC 27709 (US). **HAMMER, Philip, Eugene** [US/US]; 212A Waldo Street, Cary, NC 27511 (US). **HILL, Dwight, Steven** [US/US]; Syngenta Biotechnology, Inc., 3054 Cornwallis Road, Research Triangle Park, NC 27709 (US). **PACHLATKO, Johannes, Paul** [CH/CH]; Syngenta Crop Protection AG, Schwarzwaldallee 215, CH-4058 Basel (CH). **BUCKEL, Thomas, Günter** [DE/CH]; Syngenta Crop Protection AG, Schwarzwallee 215, CH-4058 Basel (CH).

(74) Agent: **BASTIAN, Werner**; c/o Syngenta Participations AG, Intellectual Property, P.O. Box, CH-4002 Basel (CH).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/092801 A2

(54) Title: METHODS AND COMPOSITIONS FOR MAKING EMAMECTIN

(57) Abstract: Disclosed is a family of P450 monooxygenases, each member of which regioselectively oxidizes avermectin to 4"-keto-avermectin. The P450 monooxgenases find use in methods and formulations for making emamectin from avermectin. Also disclosed are methods for purifying the P450 monooxygenases of the invention, binding agents that specifically bind to the P450 monooxygenases of the invention, and genetically engineered cells that express the P450 monooxygenases of the invention. Also disclosed are ferrodoxins and ferredoxin reductases that are active with the P450 monooxygenases of the invention.

METHODS AND COMPOSITIONS FOR MAKING EMAMECTIN

The invention relates to the field of agrochemicals, and in particular, to insecticides. More specifically, this invention relates to the derivatization of avermectin, particularly for the synthesis of emamectin.

Emamectin is a potent insecticide and controls many pests such as thrips, leafminers, and worm pests including alfalfa caterpillar, beet armyworm, cabbage looper, corn earworm, cutworms, diamondback moth, tobacco budworm, tomato fruitworm, and tomato pinworm. Emamectin (4"-deoxy-4"-epi-N-methylamino avermectin Bla/Blb) is described in U.S. Patent No. 4,874,749 and in Cvetovich, R.J. *et al.*, *J. Organic Chem.* **59**:7704-7708, 1994 (as MK-244).

U.S. Patent No. 5,288,710 describes salts of emamectin that are especially valuable agrochemically. These salts of emamectin are valuable pesticides, especially for combating insects and representatives of the order Acarina. Some pests for which emamectin is useful in combating are listed in European Patent Application EP-A 736,252.

One drawback to the use of emamectin is the difficulty of its synthesis from avermectin. This is due to the first step of the process, which is the most costly and time-consuming step of producing emamectin, in which the 4"-carbinol group of avermectin must be oxidized to a ketone. The oxidation of the 4"-carbinol group is problematic due to the presence of two other hydroxyl groups on the molecule that must be chemically protected before oxidation and deprotected after oxidation. Thus, this first step, significantly increases the overall cost and time of producing emamectin from avermectin.

Because of the efficacy and potency of emamectin as an insecticide, there is a need to develop a cost and time effective method and/or reagent for regioselectively oxidizing the 4"-carbinol group of avermectin to produce 4"-keto-avermectin, which is a necessary intermediate for producing emamectin from avermectin.

The invention provides a novel family of P450 monooxygenases, each member of which is able to regioselectively oxidize the 4"-carbinol group of unprotected avermectin, thereby resulting in a cheap, effective method to produce 4"-keto-avermectin, a necessary intermediate in the production of emamectin. The invention allows elimination of the costly, time-consuming steps of (1) chemically protecting the two other hydroxyl groups on the avermectin

molecule prior to oxidation of the 4"-carbinol group that must be chemically protected before oxidation; and (2) chemically deprotecting these two other hydroxyl groups after oxidation. The invention thus provides reagents and methods for significantly reducing the overall cost of producing emamectin from avermectin.

- Accordingly, in one aspect, the invention provides a purified nucleic acid molecule encoding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.
- In a specific embodiment , the invention relates to an purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin, which polypeptide is substantially similar, and preferably has between at least 50%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO:2, with each individual number within this range of between 50% and 99% also being part of the invention.
- In a further specific embodiment , the invention relates to an purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO:2.
- The invention further provides a purified nucleic acid molecule comprising a nucleotide sequence
  a) as given in SEQ ID NO:1;
  b) having substantial similarity to (a);
  c) capable of hybridizing to (a) or the complement thereof;
  d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO:1, or the complement thereof;
  e) complementary to (a), (b) or (c);
  f) which is the reverse complement of (a), (b) or (c), or

g) which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.

- In a specific embodiment , the invention relates to a purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin, which polypeptide is substantially similar, and preferably has at least between 60%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95, with each individual number within this range of between 60% and 99% also being part of the invention.
- In a further specific embodiment , the invention relates to an purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.
- The invention further provides a purified nucleic acid molecule comprising a nucleotide sequence

a) as given in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94;

b) having substantial similarity to (a);

c) capable of hybridizing to (a) or the complement thereof;

d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO:1, SEQ ID

NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94 or the complement thereof;

e) complementary to (a), (b) or (c);

f) which is the reverse complement of (a), (b) or (c); or

g) which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.

In certain embodiments, the nucleic acid molecule comprises or consists essentially of a nucleic acid sequence that is at least between 66%, and 99% identical to SEQ ID NO:1, with each individual number within this range of between 66%, and 99% also being part of the invention..

- In certain embodiments, the nucleic acid molecule comprises or consists essentially of a nucleic acid sequence that is at least between 70%, and 99% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94, with each individual number within this range of between 70%, and 99% also being part of the invention..
- In some embodiments, the nucleic acid molecule comprises or consists essentially of a nucleic acid sequence that is at least 80% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94.
- In certain embodiments, the nucleic acid molecule comprises or consists essentially of a nucleic acid sequence that is at least 90% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94.

- In certain embodiments, the nucleic acid molecule comprises or consists essentially of a nucleic acid sequence that is at least 95% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94.
- In some embodiments, the nucleic acid molecule comprises or consists essentially of a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, and SEQ ID NO:94.
- In particular embodiments, the nucleic acid molecule is isolated from a *Streptomyces* strain. In certain embodiments, the *Streptomyces* strain is selected from the group consisting of *Streptomyces tubercidicus*, *Streptomyces lydicus*, *Streptomyces platensis*, *Streptomyces chattanoogensis*, *Streptomyces kasugaensis*, and *Streptomyces rimosus* and *Streptomyces albofaciens*..
- In some embodiments of this aspect, the nucleic acid molecule further comprises a nucleic acid sequence encoding a tag which is linked to the P450 monooxygenase via a covalent bond. In certain embodiments, the tag is selected from the group consisting of a His tag, a GST tag, an HA tag, a HSV tag, a Myc-tag, and VSV-G-Tag.
- In another aspect, the invention provides a purified polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.
- In some embodiments, the polypeptide comprises or consists essentially of an amino acid sequence that is encoded by a nucleic acid molecule
  a) as given in SEQ ID NO:1 or the complement thereof;
  b) having substantial similarity to (a);
  c) capable of hybridizing to (a) or the complement thereof;
  d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO:1, or the complement thereof;
  e) complementary to (a), (b) or (c);

f) which is the reverse complement of (a), (b) or (c); or.

g) which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.

In some embodiments, the polypeptide comprises or consists essentially of an amino acid sequence that is between at least 50%, and 99% identical to SEQ ID NO:2, with each individual number within this range of between 50% and 99% also being part of the invention..

- In some embodiments, the polypeptide comprises or consists essentially of an amino acid sequence that is encoded by a nucleic acid molecule

a) as given in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94 or the complement thereof;

b) having substantial similarity to (a);

c) capable of hybridizing to (a) or the complement thereof;

d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94 or the complement thereof, or the complement thereof;

e) complementary to (a), (b) or (c);

f) which is the reverse complement of (a), (b) or (c); or

g) which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.

- In some embodiments, the P450 monooxygenase comprises or consists essentially of an amino acid sequence that is between at least 60%, and 99% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24,

SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95, with each individual number within this range of between 60% and 99% also being part of the invention..

- In certain embodiments, the P450 monooxygenase comprises or consists essentially of an amino acid sequence that is at least 70% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.
- In some embodiments, the P450 monooxygenase comprises or consists essentially of an amino acid sequence that is at least 80% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.
- In some embodiments, the P450 monooxygenase comprises or consists essentially of an amino acid sequence that is at least 90% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.
- In certain embodiments, the P450 monooxygenase comprises or consists essentially of an amino acid sequence that is at least 95% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.
- In some embodiments of this aspect of the invention, the P450 monooxygenase comprises or consists essentially of an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, and SEQ ID NO:95.
- In certain embodiments, the polypeptide according to the invention exhibiting an enzymatic activity of a P450 monooxygenase further comprises a tag. In some

embodiments, the tag is selected from the group consisting of a His tag, a GST tag, an HA tag, a HSV tag, a Myc-tag, and VSV-G-Tag.

- In another aspect, the invention provides a binding agent that specifically binds to a polypeptide according to the invention exhibiting an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin. In some embodiments, the binding agent is an antibody. In certain embodiments, the antibody is a polyclonal antibody or a monoclonal antibody.
- In yet another aspect, the invention provides a family of P450 monooxygenase polypeptides, wherein each member of the family regioselectively oxidizes avermectin to 4"-keto-avermectin.
- In certain embodiments, each member of the family comprises or consists essentially of an amino acid sequence that is between at least 50%, and 99% identical to SEQ ID NO:2, with each individual number within this range of between 50% and 99% also being part of the invention..
- In certain embodiments, each member of the family comprises or consists essentially of an amino acid sequence that is between at least 60%, and 99% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95, with each individual number within this range of between 60% and 99% also being part of the invention..
- In some embodiments, each member of the family comprises or consists essentially of an amino acid sequence that is at least 70% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.
- In certain embodiments, each member of the family comprises or consists essentially of an amino acid sequence that is at least 80% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95. In

some embodiments, each member of the family comprises or consists essentially of an amino acid sequence that is at least 90% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.

- In certain embodiments, each member of the family comprises or consists essentially of an amino acid sequence that is at least 95% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.
- In some embodiments of this aspect of the invention, each member of the family comprises or consists essentially of an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, and SEQ ID NO:95.
- In still another aspect, the invention provides a cell genetically engineered to comprise a nucleic acid molecule encoding a polypeptide which exhibits an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin.
- In some embodiments, the nucleic acid molecule is positioned for expression in the cell. In certain embodiments, the cell further comprises a nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin protein.
- In some embodiments, the cell further comprises a nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin reductase protein.
- In certain embodiments, the cell is a genetically engineered *Streptomyces* strain. In certain embodiments, the cell is a genetically engineered *Streptomyces lividans* strain. In particular embodiments, the genetically engineered *Streptomyces lividans* strain has NRRL Designation No. B-30478. In some embodiments, the cell is a genetically engineered *Pseudomonas* strain. In some embodiments, the cell is a genetically engineered

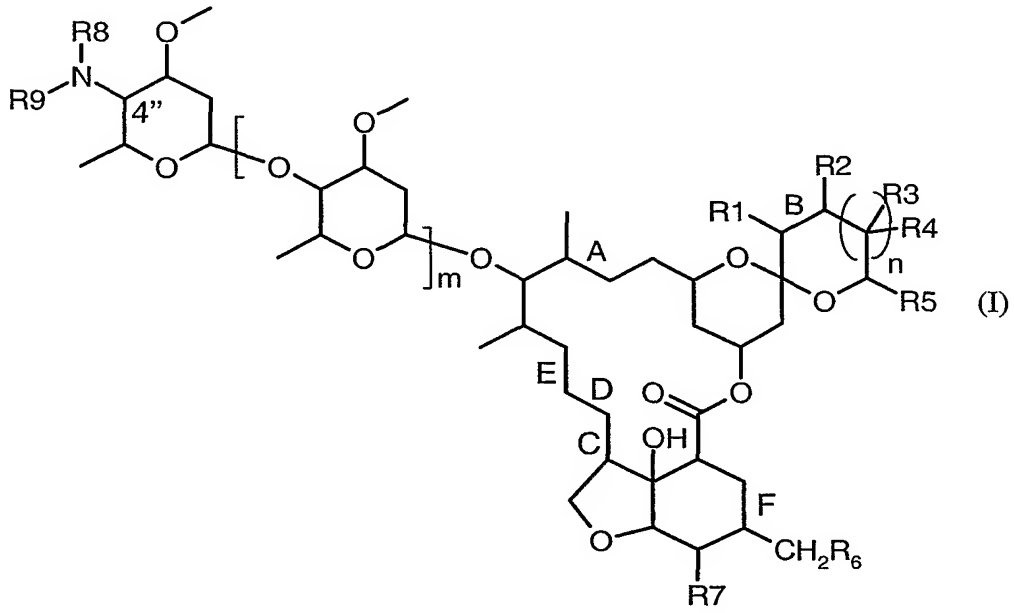
Pseudomonas putida strain. In certain embodiments, the genetically engineered *Pseudomonas putida* strain has NRRL Designation No. B-30479. In some embodiments, the cell is a genetically engineered *Escherichia coli* strain.

- In another aspect, the invention provides a purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin, wherein the nucleic acid molecule is isolated from a *Streptomyces* strain comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin.
- In a specific embodiment , the invention relates to an purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits the enzymatic activity of a ferredoxin, which polypeptide is substantially similar, and preferably has between at least 80%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO:36 or SEQ ID NO: 38, with each individual number within this range of between 80% and 99% also being part of the invention.
- In still a further specific embodiment , the invention relates to an purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits the enzymatic activity of a ferredoxin, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 36 or SEQ ID NO: 38.
- The invention further provides a purified nucleic acid molecule comprising a nucleotide sequence
  a) as given in SEQ ID NO:35 or SEQ ID NO: 37;
  b) having substantial similarity to (a);
  c) capable of hybridizing to (a) or the complement thereof;
  d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO: 35 or SEQ ID NO: 37, or the complement thereof;
  e) complementary to (a), (b) or (c);
  f) which is the reverse complement of (a), (b) or (c); or
  g) which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a ferredoxin and regioselectively oxidizes avermectin to 4"-keto-avermectin.

- In certain embodiments, the nucleic acid molecule encoding a ferredoxin of the invention comprises or consists essentially of a nucleic acid sequence that is at least 81% identical to SEQ ID NO:35 or SEQ ID NO:37. In some embodiments, the nucleic acid molecule comprises or consists essentially of a nucleic acid sequence that is at least 85%, or at least 90%, or at least 95%, or at least 99% identical to SEQ ID NO:35 or SEQ ID NO:37. In certain embodiments, the nucleic acid molecule encoding a ferredoxin of the invention comprises or consists essentially of the nucleic acid sequence of SEQ ID NO:35 or SEQ ID NO:37.
- In yet another aspect, the invention provides a purified ferredoxin protein, wherein the ferredoxin protein is isolated from a *Streptomyces* strain comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin. In certain embodiments, the ferredoxin of the invention comprises or consists essentially of an amino acid sequence that is at least 80% identical to SEQ ID NO:36 or SEQ ID NO:38. In some embodiments, the nucleic acid molecule comprises or consists essentially of an amino acid sequence that is at least 85%, or at least 90%, or at least 95%, or at least 99% identical to SEQ ID NO:36 or SEQ ID NO:38.
- In particular embodiments, the ferredoxin of the invention comprises or consists essentially of the amino acid sequence of SEQ ID NO:36 or SEQ ID NO:38.
- In another aspect, the invention provides a purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin reductase, wherein the nucleic acid molecule is isolated from a *Streptomyces* strain comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin.
- In certain embodiments, the nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin reductase comprises or consists essentially of the nucleic acid sequence of SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:102, or SEQ ID NO:104.
- In yet another aspect, the invention provides a purified polypeptide exhibiting an enzymatic activity of a ferredoxin reductase protein, wherein the said polypeptide is isolated from a *Streptomyces* strain comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin. In certain embodiments, the

polypeptide of the invention comprises or consists essentially of the amino acid sequence of SEQ ID NO:99, SEQ ID NO:101, SEQ ID NO:103, or SEQ ID NO:105.

- In another aspect, the invention provides a process for the preparation a compound of the formula



(I)

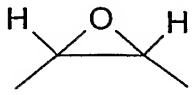
in which

R_1-R_9 represent, independently of each other hydrogen or a substituent;

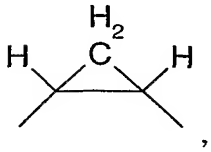
m is 0, 1 or 2;

n is 0, 1, 2 or 3; and

the bonds marked with A, B, C, D, E and F indicate, independently of each other, that two adjacent carbon atoms are connected by a double bond, a single bond, a single bond and a epoxide bridge of the formula



, or a single bond and a methylene bridge of the formula

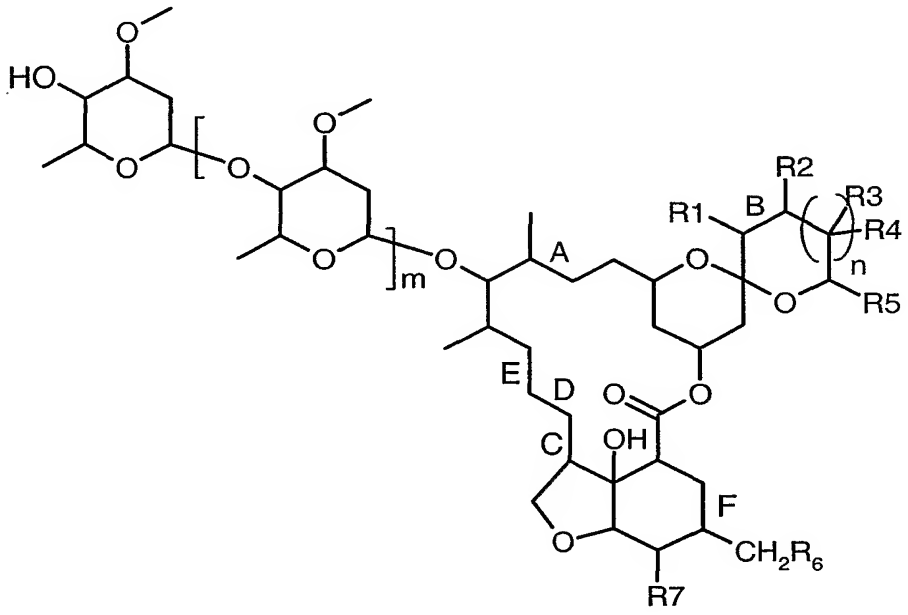


,

including, where applicable, an E/Z isomer, a mixture of E/Z isomers, and/or a tautomer thereof, in each case in free form or in salt form,

which process comprises

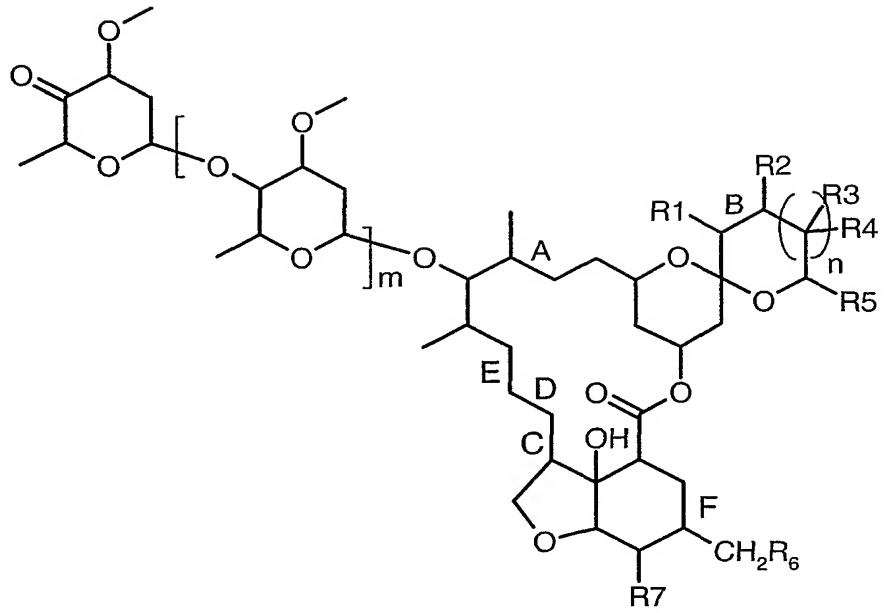
1) bringing a compound of the formula



(II);

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (I) above, into contact with a polypeptide according to the invention that is capable of regioselectively oxidising the alcohol at position 4" in order to form a compound of the formula



(III),

in which R_1, R_2, R_3, R_4, R_5, R_6, R_7, m, n, A, B, C, D, E and F have the meanings given for formula (I); and

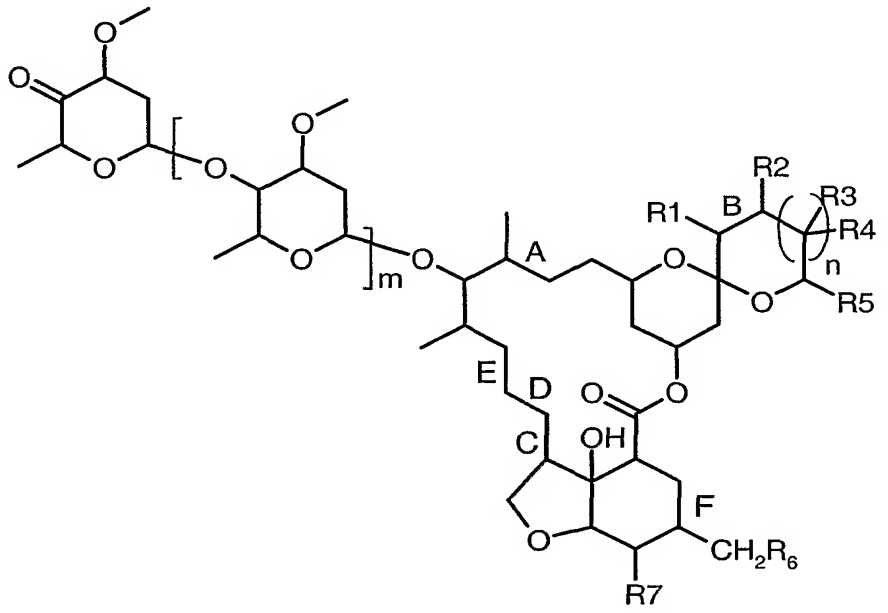
2) reacting the compound of the formula (III) with an amine of the formula $HN(R_8)R_9$, wherein R_8 and R_9 have the same meanings as given for formula (I), and which is known, in the presence of a reducing agent;

and, in each case, if desired, converting a compound of formula (I) obtainable in accordance with the process or by another method, or an E/Z isomer or tautomer thereof, in each case in free form or in salt form, into a different compound of formula (I) or an E/Z isomer or tautomer thereof, in each case in free form or in salt form, separating a mixture of E/Z isomers obtainable in accordance with the process and isolating the desired isomer, and/or converting a free compound of formula (I) obtainable in accordance with the process or by another method, or an E/Z isomer or tautomer thereof, into a salt or converting a salt, obtainable in accordance with the process or by another method, of a compound of formula (I) or of an E/Z isomer or tautomer thereof into the free compound of formula (I) or an E/Z isomer or tautomer thereof or into a different salt.

- In some embodiments, the compound of formula (II) is further brought into contact with a polypeptide according to the invention exhibiting an enzymatic activity of a

ferredoxin. In certain embodiments, the compound of formula (II) is further brought into contact with a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin reductase. In some embodiments, the compound of formula (II) is further brought into contact with a reducing agent (*e.g.*, NADH or NADPH).

- In still a further embodiment, the invention provides a process for the preparation of a compound of the formula

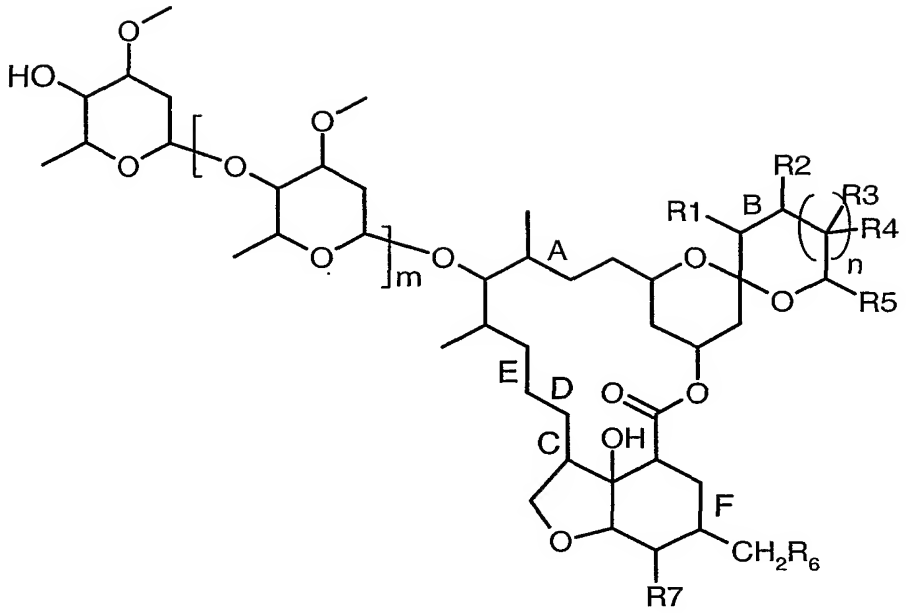


(III),

in which R_1, R_2, R_3, R_4, R_5, R_6, R_7, m, n, A, B, C, D, E and F have the meanings given for formula (I) of claim 1,

which process comprises

1) bringing a compound of the formula



(II);

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (I) above, into contact with a polypeptide according to the invention that is capable of regioselectively oxidising the alcohol at position 4", maintaining said contact for a time sufficient for the oxidation reaction to occur and isolating and purifying the compound of formula (II).

- In yet another embodiment, the invention provides a process according to the invention for the preparation of a compound of the formula (I), in which

  n is 1;

  m is 1;

  A is a double bond;

  B is single bond or a double bond,

  C is a double bond,

  D is a single bond,

  E is a double bond,

  F is a double bond; or a single bond and a epoxy bridge; or a single bond and a methylene bridge;

  R_1, R_2 and R_3 are H;

R_4 is methyl;

R_5 is C_1-C_{10}-alkyl, C_3-C_8-cycloalkyl or C_2-C_{10}-alkenyl;

R_6 is H;

R_7 is OH;

R_8 and R_9 are independently of each other H; C_1-C_{10}-alkyl or C_1-C_{10}-acyl; or together form $-(CH_2)_q-$; and

q is 4, 5 or 6.

- In still another embodiment, the invention provides a process according to the invention for the preparation of a compound of the formula (I), in which

n is 1;

m is 1;

A, B, C, E and F are double bonds;

D is a single bond;

R_1, R_2, and R_3 are H;

R_4 is methyl;

R_5 is s-butyl or isopropyl;

R_6 is H;

R_7 is OH;

R_8 is methyl

R_9 is H.

- In still another embodiment, the invention provides a process according to the invention for the preparation of 4''-deoxy-4''-N-methylamino avermectin B_{1a}/B_{1b} benzoate salt.
- In another aspect, the invention provides a method for making emamectin. The method comprises adding a polypeptide according to the invention exhibiting an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin to a reaction mixture comprising avermectin and incubating the reaction mixture under conditions that allow the polypeptide to regioselectively oxidize avermectin to 4"-keto-avermectin. In some embodiments, the reaction mixture further comprises a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin. In

certain embodiments, the reaction mixture further comprises a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin reductase. In some embodiments, the reaction mixture further comprises a reducing agent (*e.g.*, NADH or NADPH).

- In still another aspect, the invention provides a formulation for making a compound of formula (I) comprising a polypeptide according to the invention exhibiting a P450 monooxygenase activity that is capable of regioselectively oxidising the alcohol at position 4" in order to form a compound of formula (II). In some embodiments, the formulation further comprises a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin (*e.g.*, a ferredoxin from cell or strain from which the P450 monooxygenase was isolated or derived).
- In still another aspect, the invention provides a formulation for making emamectin comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin. In some embodiments, the formulation further comprises a ferredoxin (*e.g.*, a ferredoxin from cell or strain from which the P450 monooxygenase was isolated or derived).
- In certain embodiments, the formulation further comprises a polypeptide according to the invention exhibiting an enzymatic activity of a ferredoxin reductase (*e.g.*, a ferredoxin from cell or strain from which the P450 monooxygenase was isolated or derived). In some embodiments, the formulation further comprises a reducing agent (*e.g.*, NADH or NADPH).

Brief Description of the Drawings

Figure 1 is a diagrammatic representation showing a map of plasmid pTBBKA. Recognition sites by the indicated restriction endonucleases are shown, along with the location of the site in the nucleotide sequence of the plasmid. Also shown are genes (*e.g.*, kanamycin resistance "KanR"), and other functional aspects (*e.g.*, Tip promoter) contained in the plasmid.

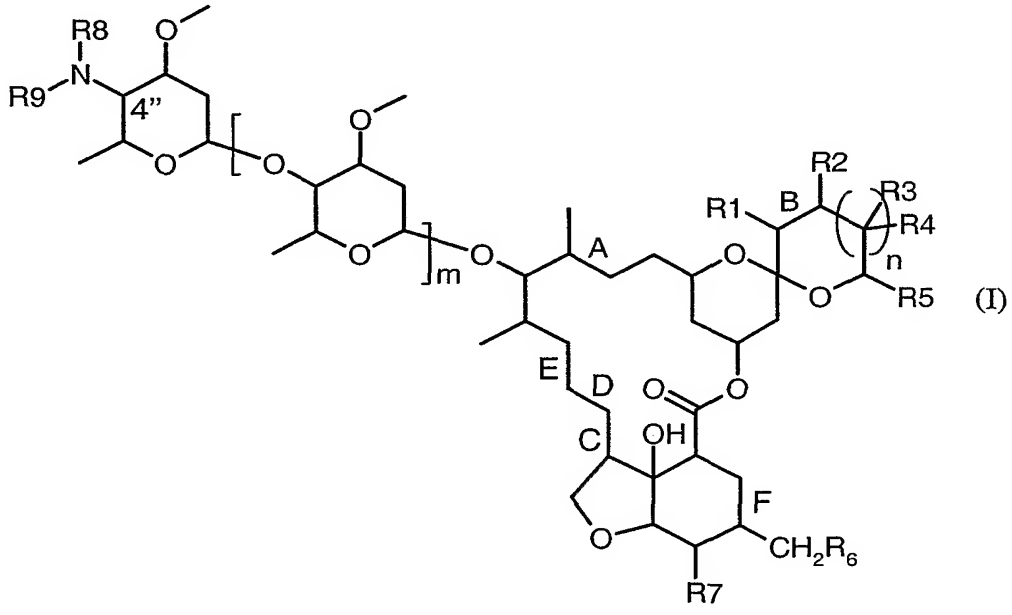
Figure 2 is a diagrammatic representation showing a map of plasmid pTUA1A. Recognition sites by the indicated restriction endonucleases are shown, along with the

location of the site in the nucleotide sequence of the plasmid. Also shown are genes (*e.g.*, ampicillin resistance "AmpR") and other functional aspects (*e.g.*, Tip promoter) contained in the plasmid.

Figure 3 is a diagrammatic representation showing a map of plasmid pRK-*ema1/fd233*. This plasmid was derived by ligating a BglII fragment containing the *ema1* and *fd233* genes organized on a single transcriptional unit into the BglII site of the broad host-range plasmid pRK290. The *ema1/fd233* genes are expressed by the tac promoter (Ptac), and they are followed by the tac terminator (Ttac). Restriction endonuclease recognition sites shown are BglII "B"; EcoRI "E"; PacI "Pc"; PmeI "Pm"; and SalI "S."

The present invention provides a family of polypeptides which exhibit an enzymatic activity of a P450 monooxygenases and are capable of regioselectively oxidizing the alcohol at position 4" of a compound of formular (II) such as avermectin in order to produce a compound of the formula (III), but especially 4"-keto-avermectin.

More particularly, the family of polypeptides according to the invention may be used in a process for the preparation a compound of the formula



(I)

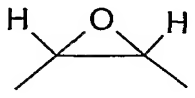
in which

R_1-R_9 represent, independently of each other hydrogen or a substituent;

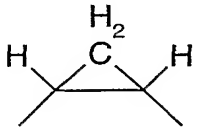
m is 0, 1 or 2;

n is 0, 1, 2 or 3; and

the bonds marked with A, B, C, D, E and F indicate, independently of each other, that two adjacent carbon atoms are connected by a double bond, a single bond, a single bond and a epoxide bridge of the formula



, or a single bond and a methylene bridge of the formula

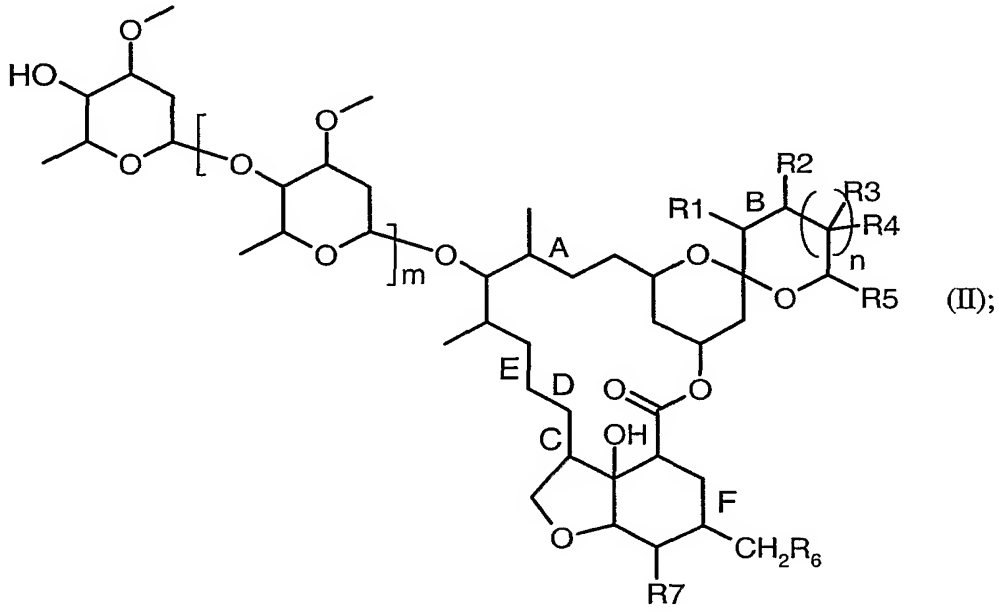


,

including, where applicable, an E/Z isomer, a mixture of E/Z isomers, and/or a tautomer thereof, in each case in free form or in salt form,

which process comprises

1) bringing a compound of the formula

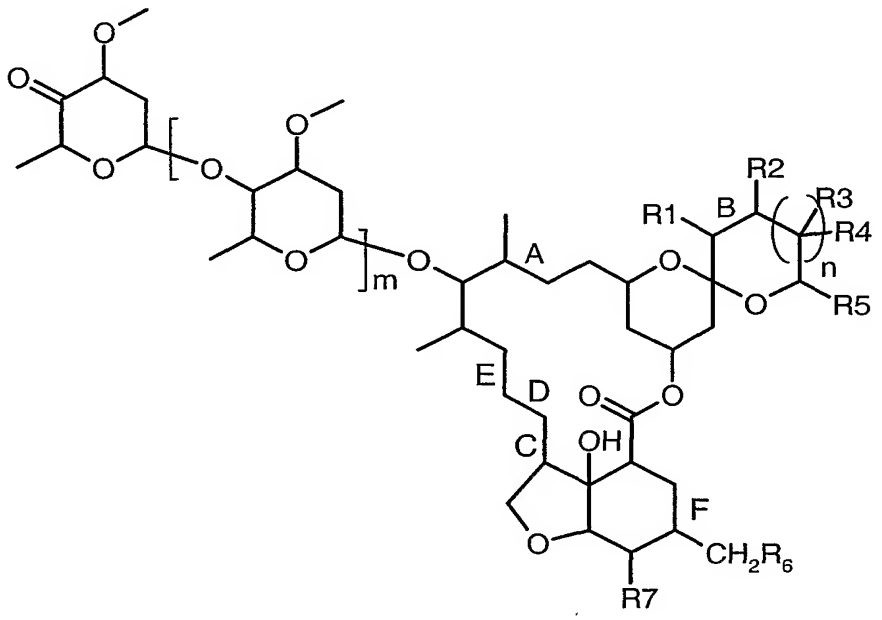


(II);

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (I) above,

into contact with a polypeptide according to the invention which exhibits an enzymatic activity of a P450 monooxygenases and is capable of regioselectively oxidizing the alcohol at position 4" of formular (II) in order to produce a compound of the formula (III)



(III),

in which R_1, R_2, R_3, R_4, R_5, R_6, R_7, m, n, A, B, C, D, E and F have the meanings given for formula (I); and

2) reacting the compound of the formula (III) with an amine of the formula $HN(R_8)R_9$, wherein R_8 and R_9 have the same meanings as given for formula (I), and which is known, in the presence of a reducing agent;

and, in each case, if desired, converting a compound of formula (I) obtainable in accordance with the process or by another method, or an E/Z isomer or tautomer thereof, in each case in free form or in salt form, into a different compound of formula (I) or an E/Z isomer or tautomer thereof, in each case in free form or in salt form, separating a mixture of E/Z isomers obtainable in accordance with the process and isolating the desired isomer, and/or converting a free compound of formula (I) obtainable in accordance with the process or by another method, or an E/Z isomer or tautomer thereof, into a salt or converting a salt, obtainable in accordance with the process or by another method, of a compound of formula (I) or of an E/Z isomer or tautomer thereof into the free compound of formula (I) or an E/Z isomer or tautomer thereof or into a different salt.

Methods of synthesis for the compounds of formula (I) are described in the literature. It has been found, however, that the processes known in the literature cause considerable problems during production basically on account of the low yields and the tedious procedures which have to be used. Accordingly, the known processes are not satisfactory in that respect, giving rise to the need to make available improved preparation processes for those compounds.

The compounds (I), (II) and (III) may be in the form of tautomers. Accordingly, herein-before and hereinafter, where appropriate the compounds (I), (II) and (III) are to be understood to include corresponding tautomers, even if the latter are not specifically mentioned in each case.

The compounds (I), (II) and (III) are capable of forming acid addition salts. Those salts are formed, for example, with strong inorganic acids, such as mineral acids, for example perchloric acid, sulfuric acid, nitric acid, nitrous acid, a phosphoric acid or a hydrohalic acid, with strong organic carboxylic acids, such as unsubstituted or substituted, for example halo-substituted, C_1-C_4alkanecarboxylic acids, for example acetic acid, saturated or unsaturated dicarboxylic acids, for example oxalic, malonic, succinic, maleic, fumaric or phthalic acid, hydroxycarboxylic acids, for example ascorbic, lactic, malic, tartaric or citric acid, or benzoic acid, or with organic sulfonic acids, such as unsubstituted or substituted, for example halo-substituted, C_1-C_4alkane- or aryl-sulfonic acids, for example methane- or p-toluene-sulfonic acid. Furthermore, compounds of formula (I), (II) and (III) having at least one acidic group are capable of forming salts with bases. Suitable salts with bases are, for example, metal salts, such as alkali metal or alkaline earth metal salts, for example sodium, potassium or magnesium salts, or salts with ammonia or an organic amine, such as morpholine, piperidine, pyrrolidine, a mono-, di- or tri-lower alkylamine, for example ethyl-, diethyl-, triethyl- or dimethyl-propyl-amine, or a mono-, di- or tri-hydroxy-lower alkylamine, for example mono-, di- or tri-ethanolamine. In addition, corresponding internal salts may also be formed. Preference is given within the scope of the invention to agrochemically advantageous salts. In view of the close relationship between the compounds of formula (I), (II) and (III) in free form and in the form of their salts, any reference hereinbefore or hereinafter to the free compounds of formula (I), (II) and (III) or to their respective salts is to be understood as including also the corresponding salts or the free compounds of formula (I), (II) and (III), where appropriate and

expedient. The same applies in the case of tautomers of compounds of formula (I), (II) and (III) and the salts thereof. The free form is generally preferred in each case.

Preferred within the scope of this invention is a process for the preparation of compounds of the formula (I), in which

n is 1;

m is 1;

A is a double bond;

B is single bond or a double bond,

C is a double bond,

D is a single bond,

E is a double bond,

F is a double bond; or a single bond and a epoxiy bridge; or a single bond and a methylene bridge;

R_1, R_2 and R_3 are H;

R_4 is methyl;

R_5 is C_1-C_{10}-alkyl, C_3-C_8-cycloalkyl or C_2-C_{10}-alkenyl;

R_6 is H;

R_7 is OH;

R_8 and R_9 are independently of each other H; C_1-C_{10}-alkyl or C_1-C_{10}-acyl; or together form $-(CH_2)_q-$; and

q is 4, 5 or 6.

Especially preferred within the scope of this invention is a process for the preparation of a compound of the formula (I) in which

n is 1;

m is 1;

A, B, C, E and F are double bonds;

D is a single bond;

R_1, R_2, and R_3 are H;

R_4 is methyl;

R_5 is s-butyl or isopropyl;

R_6 is H;

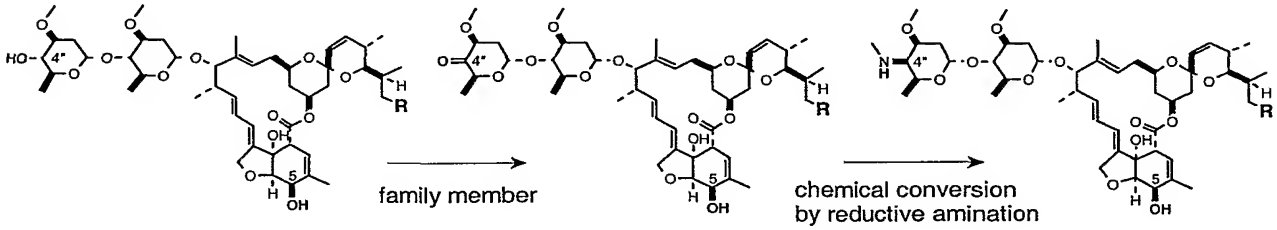
R_7 is OH;

R_8 is methyl

R_9 is H.

Very especially preferred is a process for the preparation of emamectin, more particularly the benzoate salt of emamectin. Emamectin is a mixture of 4''-deoxy-4''-N-methylamino avermectin B_{1a}/B_{1b} and is described in US-P-4,4874,749 and as MK-244 in Journal of Organic Chemistry, Vol. 59 (1994), 7704-7708. Salts of emamectin that are especially valuable agrochemically are described in US-P-5,288,710. Each member of this family of peptides exhibiting an enzymatic activity of a P450 monooxygenases as described hereinbefore is able to oxidize unprotected avermectin regioselectively at position 4", thus opening a new and more economical route for the production of emamectin.

The family members each catalyze the following reaction:



avermectin	*4"-keto-avermectin*	*emamectin*
B1a (R=CH_3) and B1b (R=H)	*B1a (R=CH_3) and B1b (R=H)*	*R=CH_3,H*

Accordingly, the invention provides a purified nucleic acid molecule encoding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and is capable of regioselectively oxidizing the alcohol at position 4" of a compound of formular (II) such as avermectin in order to produce a compound of formula (III), but especially 4"-keto-avermectin.

In particular, the invention provides a purified nucleic acid molecule encoding a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin. A "nucleic

acid molecule" refers to single-stranded or double-stranded polynucleotides, such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or analogs of either DNA or RNA.

The invention also provides a purified polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and is capable of regioselectively oxidizing the alcohol at position 4" of a compound of formular (II) such as avermectin in order to produce a compound of formula (III), but especially 4"-keto-avermectin.

In particular, the invention also provides a purified P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin.

As used herein, by "purified" is meant a nucleic acid molecule or polypeptide (*e.g.*, an enzyme or antibody) that has been separated from components which naturally accompany it. An example of such a nucleotide sequence or segment of interest "purified" from a source, would be nucleotide sequence or segment that is excised or removed from said source by chemical means, e.g., by the use of restriction endonucleases, so that it can be further manipulated, e.g., amplified, for use in the invention, by the methodology of genetic engineering. Such a nucleotide sequence or segment is commonly referred to as "recombinant.". In one specific aspect, the purified nucleic acid molecule may be separated from nucleotide sequences, such as promoter or enhancer sequences, that flank the nucleic acid molecule as it naturally occurs in the chromosome.

In the case of a protein or a polypeptide, the purified protein and polypeptide, respectively, is separated from components, such as other proteins or fragments of cell membrane, that accompany it in the cell. Of course, those of ordinary skill in molecular biology will understand that water, buffers, and other small molecules may additionally be present in a purified nucleic acid molecule or purified protein preparation. A purified nucleic acid molecule or purified polypeptide (*e.g.*, enzyme) of the invention that is at least 95% by weight, or at least 98% by weight, or at least 99% by weight, or 100% by weight free of components which naturally accompany the nucleic acid molecule or polypeptide.

According to the invention, a purified nucleic acid molecule may be generated, for example, by excising the nucleic acid molecule from the chromosome. It may then be ligated into an expression plasmid. Other methods for generating a purified nucleic acid molecule encoding a P450 monooxygenase of the invention are available and include, without limitation, artificial synthesis of the nucleic acid molecule on a nucleic acid synthesizer.

Similarly, a purified P450 monooxygenase of the invention may be generated, for example, by recombinant expression of a nucleic acid molecule encoding the P450 monooxygenase in a cell in which the P450 monooxygenase does not naturally occur. Of course, other methods for obtaining a purified P450 monooxygenase of the invention include, without limitation, artificial synthesis of the P450 monooxygenase on a polypeptide synthesizer and isolation of the P450 monooxygenase from a cell in which it naturally occurs using, *e.g.*, an antibody that specifically binds the P450 monooxygenase.

Biotransformations of secondary alcohols to ketones by *Streptomyces* bacteria are known to be catalyzed by dehydrogenases or oxidases. However, prior to the present discovery of the cytochrome P450 monooxygenase from *Streptomyces tubercidicus* strain R-922 responsible for the regioselective oxidation of avermectin to 4"-keto-avermectin, no experimental data of another cytochrome P450 monooxygenase from *Streptomyces* to oxidize a secondary alcohol to a ketone had been reported.

According to some embodiments of the invention, the nucleic acid molecule and/or the polypeptide encoded by the nucleic acid molecule are isolated from a *Streptomyces* strain. Thus, the nucleic acid molecule (or polypeptide encoded thereby) may be isolated from, without limitation, *Streptomyces tubercidicus*, *Streptomyces lydicus, Streptomyces platensis, Streptomyces chattanoogensis*, *Streptomyces kasugaensis*, *Streptomyces rimosus,* and *Streptomyces albofaciens.*

As mentioned above and described in more detail below, an entire family of polypeptides exhibiting an enzymatic activity of P450 monooxygenases capable of regioselectively oxidizing avermectin to 4"-keto-avermectin are provided herein. All of these family members are related by at least 60% identity at the amino acid level. A useful nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide of the invention exhibiting an enzymatic activity of a P450 monooxygenase comprises or consists essentially of a nucleic acid sequence that is at least 70% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94. In certain embodiments, the nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide of the invention exhibiting an enzymatic activity of a P450 monooxygenase

comprises or consists essentially of a nucleic acid sequence that is at least 80% identical; or at least 85% identical; or at least 90% identical; or at least 95% identical; or at least 98% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94.

Similarly, the invention provides a purified polypeptide exhibiting an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin which, in some embodiments, comprises or consists essentially of an amino acid sequence that is at least 60% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95. In certain embodiments, the purified polypeptide of the invention exhibiting an enzymatic activity of a P450 monooxygenase comprises or consists essentially of an amino acid sequence that is at least 70% identical; or at least 80% identical; or at least 90% identical; or at least 95% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.

In some embodiments, the nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide of the invention exhibiting an enzymatic activity of a P450 monooxygenase comprises or consists essentially of the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94. Similarly, the purified polypeptide of the invention exhibiting an enzymatic activity of a P450 monooxygenase, in some embodiments, comprises or consists essentially of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.

To describe the sequence relationships between two or more nucleic acids or polynucleotides the following terms are used: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e) "substantial identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Preferred, non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller, 1988; the local homology algorithm of Smith et al. 1981; the homology alignment algorithm of Needleman and Wunsch 1970; the search-for-similarity-method of Pearson and Lipman 1988; the algorithm of Karlin and Altschul, 1990, modified as in Karlin and Altschul, 1993.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters. The

CLUSTAL program is well described by Higgins et al. 1988; Higgins et al. 1989; Corpet et al. 1988; Huang et al. 1992; and Pearson et al. 1994. The ALIGN program is based on the algorithm of Myers and Miller, *supra*. The BLAST programs of Altschul et al., 1990, are based on the algorithm of Karlin and Altschul *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. 1997. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships

between molecules. See Altschul et al., *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989). See http://www.ncbi.nlm.nih.gov. Alignment may also be performed manually by inspection. For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the nucleotide sequences disclosed herein is preferably made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero

and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.
(e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 66%. 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, more preferably at least 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions (see below). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two

nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a polypeptide indicates that a polypeptide comprises a sequence with at least 50%, 60%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, or even more preferably, 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (1970). An indication that two polypeptide sequences are substantially identical is that one polypeptide is immunologically reactive with antibodies raised against the second polypeptide. Thus, a polypeptide is substantially identical to a second polypeptide, for example, where the two peptides differ only by a conservative substitution.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern

hybridization are sequence dependent, and are different under different environmental parameters. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, 1984; T_m 81.5°C + 16.6 (log M) +0.41 (%GC) – 0.61 (% form) – 500/L; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point I for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point I; moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point I; low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point I. Using the equation, hybridization and wash compositions, and desired T, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T of less than 45°C (aqueous solution) or 32°C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, 1993. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point T_m for the specific sequence at a defined ionic strength and pH.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2X SSC wash at 65°C for 15 minutes (see, Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium

stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1X SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6X SSC at 40°C for 15 minutes. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, more preferably about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C and at least about 60°C for long robes (e.g., >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2X (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0. 1X SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C.

The following are examples of sets of hybridization/wash conditions that may be used to clone orthologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention: a reference nucleotide sequence preferably hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium

dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

One non-limiting source of a purified polypeptide of the invention exhibiting an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin is the cell-free extract described in the examples below. Another method for purifying a polypeptide exhibiting a P450 monooxygenase activity in accordance with the invention is to attach a tag to the protein, thereby facilitating its purification. Accordingly, the invention provides a purified polypeptide exhibiting an enzymatic activity of a P450 monooxygenase which regioselectively oxidizes avermectin to 4"-keto-avermectin, wherein the polypeptide is covalently bound to a tag. The invention further provides a nucleic acid molecule encoding such a tagged polypeptide.

As used herein, a "tag" is meant a polypeptide or other molecule covalently bound to a polypeptide of the invention, whereby a binding agent (*e.g.*, a polypeptide or molecule) specifically binds the tag. In accordance with the invention, by "specifically binds" is meant that the binding agent (*e.g.*, an antibody or Ni^{2+} resin) recognizes and binds to a particular polypeptide or chemical but does not substantially recognize or bind to other molecules in the sample. In some embodiments, a binding agent that specifically binds a ligand forms an association with that ligand with an affinity of at least $10^6\,M^{-1}$, or at least $10^7\,M^{-1}$, or at least $10^8\,M^{-1}$, or at least $10^9\,M^{-1}$ either in water, under physiological conditions, or under conditions which approximate physiological conditions with respect to ionic strength, *e.g.*, 140 mM NaCl, 5 mM $MgCl_2$. For example, a His tag is specifically bound by nickel (*e.g.*, the Ni^{2+}-charged column commercially available as His•Bind® Resin from Novagen Inc, Madison, WI). Likewise, a Myc tag is specifically bound by an antibody that specifically binds Myc.

As described below, a His tag is attached to the purified polypeptide of the invention exhibiting an enzymatic activity of a P450 monooxygenase by generating a nucleic acid molecule encoding the His-tagged polypeptide, and expressing the polypeptide in *E. coli*. These polypeptides, once expressed by *E. coli*, are readily purified by standard techniques (*e.g.*, using one of the His•Bind® Kits commercially available from Novagen or using the TALON™ Resin (and manufacturer's instructions) commercially available from Clontech Laboratories, Inc., Palo Alto, CA).

Additional tags may be attached to any or all of the polypeptides of the invention to facilitate purification. These tags include, without limitation, the HA-Tag (amino acid sequence: YPYDVPDYA (SEQ ID NO:39)), the Myc-tag (amino acid sequence: EQKLISEEDL (SEQ ID NO:40)), the HSV tag (amino acid sequence: QPELAPEDPED (SEQ ID NO:41)), and the VSV-G-Tag (amino acid sequence: YTDIEMNRLGK (SEQ ID NO:42)). Covalent attachment (*e.g.*, via a polypeptide bond) of these tags to a polypeptide of the invention allows purification of the tagged polypeptide using, respectively, an anti-HA antibody, an anti-Myc antibody, an anti-HSV antibody, or an anti-VSV-G antibody, all of which are commercially available (for example, from MBL International Corp., Watertown, MA; Novagen Inc.; Research Diagnostics Inc., Flanders, NJ).

The tagged polypeptides of the invention exhibiting a P450 monooxygenase activity may also be tagged by a covalent bond to a chemical, such as biotin, which is specifically bound by streptavidin, and thus may be purified on a streptavidin column. Similarly, the tagged P450 monooxygenases of the invention may be covalently bound (*e.g.*, via a polypeptide bond) to the constant region of an antibody. Such a tagged P450 monooxygenase may be purified, for example, on protein A sepharose.

The tagged P450 monooxygenases of the invention may also be tagged to a GST (glutathione-S-transferase) or the constant region of an immunoglobulin. For example, a nucleic acid molecule of the invention (*e.g.*, comprising SEQ ID NO:1) can be cloned into one of the pGEX plasmids commercially available from Amersham Pharmacia Biotech, Inc. (Piscataway NJ), and the plasmid expressed in *E. coli*. The resulting P450 monooxygenase encoded by the nucleic acid molecule is covalently bound to a GST (glutathione-S-transferase). These GST fusion proteins can be purified on a glutathione agarose column (commercially available from, *e.g.*, Amersham Pharmacia Biotech), and thus purified. Many of the pGEX plasmids enable easy removal of the GST portion from the fusion protein. For example, the pGEX-2T plasmid contains a thrombin recognition site between the inserted nucleic acid molecule of interest and the GST-encoding nucleic acid sequence. Similarly, the pGES-3T plasmid contains a factor Xa site. By treating the fusion protein with the appropriate enzyme, and then separating the GST portion from the P450 monooxygenase of the invention using glutathione agarose (to which the GST specifically binds), the P450 monooxygenase of the invention can be purified.

Yet another method to obtain a purified polypeptide of the invention exhibiting a P450 monooxygenase activity is to use a binding agent that specifically binds to such a polypeptide. Accordingly, the invention provides a binding agent that specifically binds to a P450 monooxygenase of the invention. This binding agent of the invention may be a chemical compound (*e.g.*, a protein), a metal ion, or a small molecule.

In particular embodiments, the binding agent is an antibody. The term "antibody" encompasses, without limitation, polyclonal antibody, monoclonal antibody, antibody fragments (*e.g.*, Fab, Fv, or Fab' fragments), single chain antibody, chimeric antibody, bi-specific antibody, antibody of any isotype (*e.g.*, IgG, IgA, and IgE), and antibody from any specifies (*e.g.*, rabbit, mouse, and human).

In one non-limiting example, the binding agent of the invention is a polyclonal antibody. In another non-limiting example, the binding agent of the invention is a monoclonal antibody. Methods for making both monoclonal and polyclonal antibodies are well known (*see, e.g.*, Current Protocols in Immunology, ed. John E. Coligan, John Wiley & Sons, Inc. 1993; Current Protocols in Molecular Biology, eds. Ausubel *et al.*, John Wiley & Sons, Inc. 2000).

The polypeptides described herein exhibiting an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin belong to a family of novel P450 monooxygenases. Accordingly, the invention also provides a family of P450 monooxygenase polypeptides, wherein each member of the family regioselectively oxidizes avermectin to 4"-keto-avermectin. In some embodiments, each member of the family comprises or consists of an amino acid sequence that is at least 50% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95. In particular embodiments, each member of the family is encoded by a nucleic acid molecule comprising or consisting of a nucleic acid sequence that is at least 66% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94.

The present invention, which provides an entire family of P450 monooxygenases, each member of which is able to regioselectively oxidize avermectin to 4"-keto-avermectin, allowed for the generation of an improved P450 monooxygenase, which may not be naturally occurring, but which regioselectively oxidizes avermectin to 4"-keto-avermectin with efficiency and with reduced undesirable side product. For instance, one of the members of the P450 monooxygenase family of the invention, $P450_{Ema1}$ enzyme catalyzes a further oxidation that is not desirable, since the formation of 3"-O-demethyl-4"-keto-avermectin has been detected in the reaction by *Streptomyces tubercidicus* strain R-922 and by *Streptomyces lividans* containing the *ema1* gene. The formation of 3"-O-demethyl-4"-keto-avermectin is brought about by the oxidation of the 3"-O-methyl group, whereby the hydrolytically labile 3"-O-hydroxymethyl group is formed which hydrolyzes to form formaldehyde and the 3"-hydroxyl group.

By providing a family of polypeptides exhibiting an enzymatic activitiy of P450 monooxygenases that regioselectively oxidize avermectin to 4"-keto-avermectin (*see*, *e.g.*, Table 3 below), individual members of the family can be subjected to family gene shuffling efforts in order to produce new hybrid genes encoding optimized P450 monooxygenases of the invention. In one non-limiting example, a portion of the *ema1* gene encoding the O_2 binding site of the $P450_{Ema1}$ protein can be swapped with the portion of the *ema2* gene encoding the O_2 binding site of the $P450_{Ema2}$ protein. Such a chimeric ema1/2 protein is within definition of a P450 monooxygenase of the invention.

Site-directed mutagenesis or directed evolution technologies may also be employed to generate derivatives of the *ema1* gene that encode enzymes with improved properties, including higher overall activity and/or reduced side product formation. One method for deriving such a mutant is to mutate the *Streptomyces* strain itself, in a manner similar to the UV mutation of *Streptomyces tubercidicus* strain R-922 described below.

Additional derivatives may be made by making conservative or non-conservative changes to the amino acid sequence of a P450 monooxygenase. Conservative and non-conservative amino acid substitutions are well known (*see*, *e.g.*, Stryer, Biochemistry, 3rd Ed., W.H. Freeman and Co., NY 1988). Similarly, truncations of a P450 monooxygenase of the invention may be generated by truncating the protein at its N-terminus (*e.g.*, see the *ema1A*

gene described below), at its C-terminus, or truncating (*i.e.*, removing amino acid residues) from the middle of the protein.

Such a mutant, derivative, or truncated P450 monooxygenase is a P450 monooxygenase of the invention as long as the mutant, derivative, or truncated P450 monooxygenase is able to regioselectively oxidize avermectin to 4"-keto-avermectin.

In another aspect, the invention provides a cell genetically engineered to comprise a nucleic acid molecule encoding a polypeptide which exhibits an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin. By "genetically engineered" is meant that the nucleic acid molecule is exogenous to the cell into which it is introduced. Introduction of the exogenous nucleic acid molecule into the genetically engineered cell may be accomplished by any means, including, without limitation, transfection, transduction, and transformation.

In certain embodiments, the nucleic acid molecule is positioned for expression in the genetically engineered cell. By "positioned for expression" is meant that the exogenous nucleic acid molecule encoding the polypeptide is linked to a regulatory sequence in such a way as to permit expression of the nucleic acid molecule when introduced into a cell. By "regulatory sequence" is meant nucleic acid sequences, such as initiation signals, polyadenylation (polyA) signals, promoters, and enhancers, which control expression of protein coding sequences with which they are operably linked. By "expression" of a nucleic acid molecule encoding a protein or polypeptide fragment is meant expression of that nucleic acid molecule as protein and/or mRNA.

A genetically engineered cell of the invention may be a prokayotic cell (*e.g.*, *E. coli*) or a eukaryotic cell (*e.g.*, *Saccharomyces cerevisiae* or mammalian cell (*e.g.*, HeLa)). According to some embodiments of the invention, the genetically engineered cell is a cell wherein the wild-type (*i.e.*, not genetically engineered) cell does not naturally contain the inserted nucleic acid molecule and does not naturally express the protein encoded by the inserted nucleic acid molecule. Accordingly, the cell may be a genetically engineered *Streptomyces* strain, such as a *Streptomyces lividans* or a Streptomyces avermitilis strain. Alternatively, the cell may be a genetically engineered *Pseudomonas* strain, such as a *Pseudomonas putida* strain or a Pseudomonas fluorescens strain. In another alternative, the cell may be a genetically engineered *Escherichia coli* strain.

Note that in some types of cells genetically engineered to comprise a nucleic acid molecule encoding a polypeptide which exhibits an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin, the actual genetically engineered cell, itself, may not be able to convert avermectin into 4"-keto-avermectin. Rather, the P450 monooxygenase heterogously expressed by such a genetically engineered cell may be purified from that cell, where the purified P450 monooxygenase of the invention can be used to regioselectively oxidize avermectin to 4"-keto-avermectin. Thus, the genetically engineered cell of the invention need not, itself, be able to regioselectively convert avermectin to 4"-keto-avermectin; rather, the genetically engineered cell of the invention need only comprise a nucleic acid molecule encoding a polypeptide which exhibits an enzymatic activity of a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin, regardless of whether the polypeptide is active inside that cell.

In addition, a cell (*e.g., E. coli*) geneticially engineered to comprise a nucleic acid molecule encoding a polypeptide of the invention which exhibits an enzymatic activity of a P450 monooxygenase may not be able to regioselectively oxidize avermectin to 4"-keto-avermectin, although the P450 monooxygenase purified from the genetically engineered cell is able to regioselectively oxidize avermectin to 4"-keto-avermectin. However, if the same cell were genetically engineered to comprise a polypeptide of the invention which exhibits an enzymatic activity of a P450 monooxygenase, a ferredoxin of the invention, and/or a ferredoxin reductase of the invention, then the P450 monooxygenase together with the ferredoxin and the ferredoxin reductase, all purified from that cell, and in the presence of a reducing agent (*e.g.*, NADH or NADPH), would be able to regioselectively oxidize avermectin to 4"-keto-avermectin. Furthermore the genetically engineered cell comprising a P450 monooxygenase of the invention, a ferredoxin of the invention, and a ferredoxin reductase of the invention, itself, might be able to carry out this oxidation.

Moreover, in a non-limiting example where a cell (*e.g., E. coli*) is genetically engineered to express P450 monooxygenase, a ferredoxin, and a ferredoxin reductase proteins of the invention, all three of these proteins, when purified from the genetically engineered *E. coli*, are together and in the presence of a reducing agent (*e.g.*, NADH or NADPH) would be active and able to regioselectively oxidize avermectin to 4"-keto-avermectin, and so are useful in a method for making emamectin.

In accordance with the present invention, the following material has been deposited with the Agricultural Research Service, Patent Culture Collection (NRRL), 1815 North University Street, Peoria, Illinois 61604, under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure: (1) *Streptomyces lividans* ZX7 (*ema1/fd233*-TUA1A) NRRL Designation No. B-30478; and (2) *Pseudomonas putida* NRRL B-4067 containing plasmid pRK290-*ema1/fd233*, NRRL Designation No.B-30479

In identifying the novel family of polypeptides exhibiting an enzymatic activity of P450 monooxygenases that regioselectively oxidize avermectin to 4"-keto-avermectin, novel ferredoxins and novel ferredoxin reductases were also identified in the same strains of bacteria in which the P450 monooxygenases were found. Accordingly, in a further aspect, the invention provides a purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits an enzymatic activity of a ferredoxin, wherein the nucleic acid molecule is isolated from a *Streptomyces* strain comprising a polypeptide that regioselectively oxidizes avermectin to 4"-keto-avermectin. Similarly, the invention provides a purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits an enzymatic activity of a ferredoxin reductase, wherein the nucleic acid molecule is isolated from a *Streptomyces* strain comprising a polypeptide that regioselectively oxidizes avermectin to 4"-keto-avermectin. The invention also provides a purified protein that exhibits an enzymatic activity of a ferredoxin, as well as a purified protein that exhibits an enzymatic activity of a ferredoxin reductase, wherein the ferredoxin protein and the ferredoxin reductase protein are isolated from a *Streptomyces* strain comprising a polypeptide that regioselectively oxidizes avermectin to 4"-keto-avermectin.

A useful nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits an enzymatic activity of a ferredoxin comprises or consists essentially of a nucleic acid sequence that is at least 81% identical to SEQ ID NO:35 or SEQ ID NO:37. Alternatively, the nucleic acid molecule comprises or consists essentially of a nucleic acid sequence that is at least 85%, or at least 90%, or at least 95%, or at least 99% identical to SEQ ID NO:35 or SEQ ID NO:37. The nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide that exhibits an enzymatic activity of a ferredoxin

may comprise or consist essentially of the nucleic acid sequence of SEQ ID NO:35 or SEQ ID NO:37.

The protein of the invention exhibiting a ferredoxin activity may comprise or consist essentially of an amino acid sequence that is at least 80% identical to SEQ ID NO:36 or SEQ ID NO:38. In some embodiments, the nucleic acid molecule comprises or consists essentially an amino acid sequence that is at least 85%, or at least 90%, or at least 95%, or at least 99% identical to SEQ ID NO:36 or SEQ ID NO:38. The ferredoxin of the invention may comprise or consist essentially of the amino acid sequence of SEQ ID NO:36 or SEQ ID NO:38.

A useful nucleic acid molecule comprising a nucleotide sequence encoding a protein of the invention exhibiting a ferredoxin reductase comprises or consists essentially of the nucleic acid sequence that is at least 85%, or at least 90%, or at least 95%, or at least 99% identical to SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:102, or SEQ ID NO:104. In a particular embodiment of the invention, the nucleic acid molecule encoding a ferredoxin reductase of the invention may comprise or consist essentially of the amino acid sequence of SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:102, or SEQ ID NO:104.

The ferredoxin reductase of the invention may comprise or consist essentially of the amino acid sequence that is at least 85%, or at least 90%, or at least 95%, or at least 99% identical to SEQ ID NO:99, SEQ ID NO:101, SEQ ID NO:103, or SEQ ID NO:105. In a particular embodiment of the invention, the ferredoxin reductase of the invention may comprise or consist essentially of the amino acid sequence of SEQ ID NO:99, SEQ ID NO:101, SEQ ID NO:103, or SEQ ID NO:105.

. Methods for purifying ferredoxin and ferredoxin reductase proteins and nucleic acid molecules encoding such ferredoxin and ferredoxin reductase proteins are known in the art and are the same as those described above for purifying P450 monooxygenases of the invention and nucleic acid molecules encoding P450 monooxygenases of the invention.

In one non-limiting example to obtain a purified P450 monooxygenase of the invention with a purified ferredoxin, a *S. lividans* strain (or *P. putida* strain, or any other cell in which the P450 monooxygenase of the invention does not naturally occur) may be genetically engineered to contain a first nucleic acid molecule encoding a P450 monooxygenase of the invention and a second nucleic acid molecule encoding a ferredoxin protein, where both the first and second nucleic acid molecules are positioned for expression in the genetically

engineered cell. The first and the second nucleic acid molecules can be on separate plasmids, or can be on the same plasmid. Thus, the same engineered cell or strain will produce both the P450 monooxygenase of the invention and the ferredoxin protein of the invention.

In a further non-limiting example to obtain a purified P450 monooxygenase of the invention with a purified ferredoxin and with a purified ferredoxin reductase of the invention, a *S. lividans* strain (or *P. putida* strain, or any other cell in which the P450 monooxygenase of the invention does not naturally occur) may be genetically engineered to contain a first nucleic acid molecule encoding a P450 monooxygenase of the invention and a second nucleic acid molecule encoding a ferredoxin protein of the invention and a third nucleic acid molecule encoding a ferredoxin reductase protein of the invention, where all the first and second and third nucleic acid molecules are positioned for expression in the genetically engineered cell. The first and the second and the third nucleic acid molecules may be provided on separate plasmids, or on the same plasmid. Thus, the same engineered cell or strain will produce all the P450 monooxygenase of the invention and the ferredoxin and the ferredoxin reductase proteins of the invention.

As described above for the P450 monooxygenases of the invention, the ferredoxin protein and/or the ferredoxin reductase protein may further comprise a tag. Moreover, the invention contemplates binding agents (*e.g.*, antibodies) that specifically bind to the ferredoxin protein, and binding agents that specifically bind to the ferredoxin reductase proteins of the invention. Methods for generating tagged ferredoxin protein, tagged ferredoxin reductase protein, and binding agents (*e.g.*, antibodies) that specifically bind to ferredoxin or ferredoxin reductase are the same as those as described above for generating tagged P450 monooxygenases of the invention and generating binding agents that specifically bind P450 monooxygenases of the invention.

The invention also provides a method for making emamectin. In this method, a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin is added to a reaction mixture containing avermectin. The reaction mixture is then incubated under conditions that allow the P450 monooxygenase to regioselectively oxidize avermectin to 4"-keto-avermectin. The reaction mixture may further comprise a ferredoxin, such as a ferredoxin of the present invention. In particular embodiments, the reaction mixture further

comprises a ferredoxin reductase such as a ferredoxin of the present invention. The reaction mixture may further comprise a reducing agent, such as NADH or NADPH.

Additionally, the invention provides a method for making 4"-keto-avermectin. The method comprises adding a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin to a reaction mixture comprising avermectin and incubating the reaction mixture under conditions that allow the P450 monooxygenase to regioselectively oxidize avermectin to 4"-keto-avermectin. In some embodiments, the reaction mixture further comprises a ferredoxin, such as a ferredoxin of the present invention. The reaction mixture may also further comprise a ferredoxin reductase such as a ferredoxin of the present invention. In particular embodiments, the reaction mixture further comprises a reducing agent, such as NADH or NADPH.

The invention also provides a formulation for making emamectin comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin. In some embodiments, the formulation further comprises a ferredoxin, such as a ferredoxin of the present invention. In particular embodiments, the ferredoxin is isolated from the same species of cell or strain from which the P450 monooxygenase was isolated or derived. The formulation may further comprise a ferredoxin reductase , such as a ferredoxin reductase of the present invention. In particular embodiments, the ferredoxin reductase is isolated from the same species of cell or strain from which the P450 monooxygenase was isolated or derived. . In some embodiments, the formulation further comprises a reducing agent, such as NADH or NADPH.

In addition, the invention provides a formulation for making 4"-keto-avermectin comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"-keto-avermectin. In some embodiments, the formulation further comprises a ferredoxin, such as a ferredoxin of the present invention. In particular embodiments, the ferredoxin is isolated from the same species of cell or strain from which the P450 monooxygenase was isolated or derived. In some embodiments, the formulation further comprises a ferredoxin reductase, such as a ferredoxin reductase of the present invention. In particular embodiments, the ferredoxin reductase is isolated from the same species of cell or strain from which the P450 monooxygenase was isolated or derived. The formulation may further comprise a reducing agent, such as NADH or NADPH.

The following examples are intended to further illustrate certain preferred embodiments of the invention and are not limiting in nature.

EXAMPLE I

Optimized Growth Conditions for *Streptomyces tubercidicus* Strain R-922

In one non-limiting example the fermentation conditions needed to provide a steady supply of cells of *Streptomyces tubercidicus* strain R-922 highly capable of regioselectively oxidizing avermectin to 4"-keto-avermectin were optimized.

First, the following solutions were made. For ISP-2 agar, 4 g of yeast extract (commercially available from Oxoid Ltd, Basingstoke, UK), 4 g of D(+)-glucose, 10 g of bacto malt extract (Difco No. 0186-17-7 (Difco products commercially available from, *e.g.*, Voigt Global Distribution, Kansas City, MO)), and 20 g of agar (Difco No. 0140-01) were dissolved in one liter of demineralized water, and the pH is adjusted to 7.0. The solution was sterilized at 121°C for 20 min., cooled down, and kept at 55°C for the time needed for the immediate preparation of the agar plates.

For PHG medium, 10 g of peptone (Sigma 0521; commercially available from Sigma Chemical Co., St. Louis, MO), 10 g of yeast extract (commercially available from Difco), 10 g of D-(+)-glucose, 2 g of NaCl, 0.15 g of $MgSO_4$ x 7 H_2O, 1.3 g of NaH_2PO_4 x H_2O, and 4.4 g of K_2HPO_4 were dissolved in 1 liter of demineralized water, and the pH was adjusted to 7.0.

Streptomyces tubercidicus strain R-922 was grown in a Petri dish on ISP-2 agar at 28°C. This culture was used to inoculate four 500 ml shaker flasks with a baffle, each containing 100 ml PHG medium. These pre-cultures were grown on an orbital shaker at 120 rpm at 28°C for 72 hours and then used to inoculate a 10-liter fermenter equipped with a mechanical stirrer and containing 8 liters of PHG medium. This main culture was grown at 28°C with stirring at 500 rpm and with aeration of 1.75 vvm (14 l/min.) and a pressure of 0.7 bar. At the end of the exponential growth, after about 20 hours, the cells were harvested by centrifugation. The yield of wet cells was 70-80 g/l culture.

EXAMPLE II

Whole Cell Biocatalysis Assay

As determined in accordance with the present invention, the following whole cell biocatalysis assay was employed to determine that the activity from *Streptomyces* cells capable of regioselectively oxidizing avermectin to 4"-keto-avermectin is catalyzed by a P450 monooxygenase.

Streptomyces tubercidicus strain R-922 was grown in PHG medium, and *Streptomyces tubercidicus* strain I-1529 was grown in M-17 or PHG medium. PHG medium contains 10 g/l Peptone (Sigma, 0.521), 10 g/l Yeast Extract (Difco, 0127-17-9), 10 g/l D-Glucose, 2 g/l NaCl, 0.15 g/l $MgSO_4$ x 7 H_2O, 1.3 g/l NaH_2PO_4 x 1 H_2O, and 4.4 g/l K_2HPO_4 at pH 7.0. M-17 medium contains 10 g/l glycerol, 20 g/l Dextrin white, 10 g/l Soytone (Difco 0437-17), 3 g/l Yeast Extract (Difco 0127-17-9), 2 g/l $(NH_4)_2SO_4$, and 2 g/l $CaCO_3$ at pH 7.0

To grow the cells, an ISP2 agar plate (not older than 1-2 weeks) was inoculated and incubated for 3-7 days until good growth was achieved. Next, an overgrown agar piece was transferred (with an inoculation loop) to a 250ml Erlenmeyer flask with 1 baffle containing 50 ml PHG medium. This pre-culture is incubated at 28°C and 120 rpm for 2-3 days. Next, 5 ml of the pre-culture were transferred to a 500 ml Erlenmeyer flask with 1 baffle containing 100 ml PHG medium. The main culture was incubated at 28°C and 120 rpm for 2 days. Next, the culture was centrifuged for 10 min. at 8000 rpm on a Beckman Rotor JA-14. The cells were next washed once with 50 mM potassium phosphate buffer, pH 7.0.

To perform the whole cell biocatalysis assay, 500 mg wet cells were placed into a 25 ml Erlenmeyer flask, to which were added 10 ml of 50 mM potassium phosphate buffer, pH 7.0. The cells were stirred with a magnetic stir bar to distribute the cells. Next, 15 µl of a solution of avermectin B1a in isopropanol (30 mg/ml) were added, and the mixture shaken on an orbital shaker at 160 rpm and 28°C. Strain R-922 was reacted for 2 hours, and strain I-1529 was reacted for 30 hours.

To work up the cultures in the whole cell biocatalysis assay, 10 ml methyl-t-butyl-ether was added to an Erlenmeyer flask containing the resting cells and the entire cell mixture was transferred to a 30 ml-centrifuge tube, shaken vigorously, and then centrifuged at 16000 rpm for 10 min. The ether phase was pipetted into a 50 ml pear flask, and evaporated in vacuo by means of a rotary evaporator (≤0.1 mbar). The residue was re-dissolved in 1.2 ml acetonitrile

and transferred to an HPLC-sample vial. The conversion of avermectin B1a to 4"-hydroxy-avermectin B1a and 4"-keto-avermectin B1a (also called 4"-oxo-avermectin B1a) and the formation of a side product from the biocatalysis reaction could be observed by HPLC analysis using HPLC protocol I.

For HPLC protocol I, the following parameters were used:

Hardware

Pump:	L-6250 Merck-Hitachi
Autosampler:	AS-2000A Merck-Hitachi
Interface Module:	D-6000 Merck-Hitachi
Channel 1-Detector:	L-7450A UV-Diode Array Merck-Hitachi
Column Oven:	none
Column:	70mm x 4mm
Adsorbent:	Kromasil 100Å-3.5µ-C18

Gradient Mode: Low

Pressure Limit:	*5-300bar*		
Column Temperature	*ambient (≈20°C)*		
Solvent A:	acetonitrile		
Solvent B:	water		
Flow:	1.5 ml/min		
Detection:	243 nm		
Pump Table:	0.0 min	75% A	25% B
linear gradient	7.0 min	100% A	0% B
	9.0 min	100% A	0% B
jump	9.1 min	75% A	25% B
	12.0 min	75% A	25% B
Stop time:	12 min		
Sampling Period:	every 200 msec		
Retention time table:	*time*	*References*	
	2.12 min	4"-hydroxy- avermectin B1a	

3.27 min avermectin B1a

3.77 min 3"-O-demethyl-4"-keto-avermectin B1a

4.83 min 4"-keto-avermectin B1a

EXAMPLE III

Biotransformation With Cell-Free Extract From *Streptomyces* Strain R-922

To prepare an active cell-free extract from *Streptomyces tubercidicus* strain R-922 capable of regioselective oxidation of avermectin to 4"-keto-avermectin, the following solutions were made, stored at 4°C, and kept on ice when used.

Solution	Formula
PP-buffer	50 mM K_2HPO_4/KH_2PO_4 (pH 7.0)
Disruption buffer	50 mM K_2HPO_4/KH_2PO_4 (pH 7.0), 5 mM benzamidine, 2 mM dithiothreitol, and 0.5 mM Pefabloc (from Roche Diagnostics)
Substrate	10 mg avermectin were dissolved in 1 ml isopropanol

Six grams of wet cells from *Streptomyces* strain R-922 were washed in PP-buffer and then resuspended in 35 ml disruption buffer and disrupted in a French press at 4°C. The resulting suspension was centrifuged for 1 hour at 35000 x g. The supernatant of the cell free extract was collected. One µl substrate was added to 499µl of cleared cell free extract and incubated at 30°C for 1 hour. Then, 1 ml methyl-t-butyl ether was added to the reaction mixture and thoroughly mixed. The mixture was next centrifuged for 2 min. at 14000 rpm, and the methyl-t-butyl ether phase was transferred into a 10 ml flask and evaporated in vacuo

by means of a rotary evaporator. The residue was dissolved in 200 µl acetonitrile and transferred into an HPLC-sample vial.

For HPLC, the HPLC protocol I was used.

When 1 µl substrate was added to 499 µl of cleared cell free extract and incubated at 30°C, no conversion of avermectin to 4"-keto-avermectin was observed by HPLC analysis using HPLC protocol I.

However, the possibility of addition of spinach ferredoxin and spinach ferredoxin reductase and NADPH to the cell free extract to restore the biocatalytic activity was explored (*see*, generally, D.E. Cane and E.I. Graziani, *J. Amer. Chem. Soc.* **120**:2682, 1998). Accordingly, the following solutions were made:

Solution	Formula
Substrate	10 mg avermectin were dissolved in 1 ml isopropanol
Ferredoxin	5 mg ferredoxin (from spinach), solution 1-3 mg/ml in Tris/HCl-buffer (from Fluka) *or* 5 mg ferredoxin (from *Clostridium pasteurianum*), solution of 1-3 mg/ml in Tris/HCl-buffer (from Fluka) *or* 5 mg ferredoxin (from *Porphyra umbilicalis*), solution of 1-3 mg/ml in Tris/HCl-buffer (from Fluka)
Ferredoxin Reductase	1 mg freeze-dried ferredoxin reductase (from spinach), solution of 3.9 U/mg in 1 ml H_2O (from Sigma)
NADPH	100 mM NADPH in H_2O (from Roche Diagnostics)

The substrate solution was stored at 4°C, the other solutions were stored at -20°C, and kept on ice when used.

Thus, to 475 µl of cleared cell free extract the following solutions were added: 10 µl ferredoxin, 10 µl ferredoxin reductase and 1 µl substrate. After the addition of substrate to the cells, the mixture was immediately and thoroughly mixed and aerated. Then, 5 µl of NADPH were added and the mixture incubated at 30°C for 30 min. Then, 1 ml methyl-t-butyl ether was added to the reaction mixture and thoroughly mixed. The mixture was next centrifuged for 2 min. at 14000 rpm, and the methyl-t-butyl ether phase was transferred into a 10 ml flask

and evaporated in vacuo by means of a rotary evaporator. The residue was dissolved in 200 µl acetonitrile and transferred into an HPLC-sample vial, and HPLC analysis performed using HPLC protocol I.

Formation of 4"-keto-avermectin was observable by HPLC analysis. Thus, addition of spinach ferredoxin and spinach ferredoxin reductase and NADPH to the cell free extract restored the biocatalytic activity.

Upon injection of a 30 µl sample, a peak appeared at 4.83 min., indicating the presence of 4"-keto-avermectin B1a. A mass of 870 D could be assigned to this peak by HPLC-mass spectrometry which corresponds to the molecular weight of 4"-keto-avermectin B1a.

Note that when analyzing product formation by HPLC and HPLC-mass spectrometry, in addition to the 4"-keto-avermectin, the corresponding ketohydrate 4"-hydroxy-avermectin was also found giving a peak at 2.12 min. This finding indicated that the P450 monooxygenase converts avermectin by hydroxylation to 4"-hydroxy-avermectin, from which 4"-keto-avermectin is formed by dehydration. Interestingly, when the spinach ferredoxin was replaced by ferredoxin from the bacterium *Clostridium pasteurianum* or from the red alga *Porphyra umbilicalis*, the biocatalytic conversion of avermectin to 4"-keto-avermectin still took place, indicating that the enzyme does not depend on a specific ferredoxin for receiving reduction equivalents.

EXAMPLE IV

Isolation of a Mutant *Streptomyces* Strain R-922 With Enhanced Activity

To obtain strains of *Streptomyces* strain R-922 that have an enhanced ability to regioselectively oxidize avermectin to 4"-keto-avermectin, UV mutants were generated. To do this, spores of *Streptomyces* strain R-922 were collected and stored in 15% glycerol at -20°C. This stock solution contained $2x10^9$ spores.

The spore stock solution was next diluted and transferred to petri plates containing 10ml of sterile water, and the suspension was exposed to UV light in a Stratalinker UV crosslinker 2400 (commercially available from Stratagene, La Jolla, CA). The Stratalinker UV crosslinker uses a 254-nm light source and the amount of energy used to irradiate a sample can be set in the "energy mode."

Applicant's or agent's file reference	International application No.
PB/5-60016A	PCT/EP 02/05363

INDICATIONS RELATING TO DEPOSITED MICROORGANISM OR OTHER BIOLOGICAL MATERIAL

(PCT Rule 13*bis*)

A. The indications made below relate to the deposited microorganism or other biological material referred to in the description on page 41, line 1 - 7.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☒

Name of depositary institution
Agricultural Research Service, Patent Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*
1815 North University Street
Peoria
Illinois 61604
USA

Date of deposit	Accession Number
May 08, 2001	NRRL B-30479

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

Pseudomonas putida NRRL B-4067 containing plasmid pRK290-ema1/fd233

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☒ This sheet was received by the International Bureau on: 02 SEP 2002
Authorized officer	Authorized officer Howard Hernandez

Form PCT/RO/134 (July1998)

Applicant's or agent's file reference	International application No.
PB/5-60016A	PCT/EP 02/05363

INDICATIONS RELATING TO DEPOSITED MICROORGANISM OR OTHER BIOLOGICAL MATERIAL

(PCT Rule 13*bis*)

A. The indications made below relate to the deposited microorganism or other biological material referred to in the description on page 41, line 1 - 7.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet [X]

Name of depositary institution
Agricultural Research Service, Patent Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*
1815 North University Street
Peoria
Illinois 61604
USA

Date of deposit	Accession Number
May 08, 2001	NRRL B-30478

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet [ ]

Streptomyces lividans ZX7 (ema1/fd233-TUA1A)

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
[ ] This sheet was received with the international application	[X] This sheet was received by the International Bureau on: 02 SEP 2002
Authorized officer	Authorized officer HUGO HERNANDEZ

Form PCT/RO/134 (July1998)

Through experimentation, it was determined that an exposure of 8000 microjoules of UV irradiation (254 nm) was required to kill 99.9% of the spores. This level of UV exposure was used in the mutagenesis.

Surviving UV-mutagenized spores were plated, cultured, and transferred to minimal media. Approximately 0.3-0.4% of the viable spores were determined to be auxotrophic, indicating a good level of mutagenesis in the population.

The mutagenized clones were screened for activity in the whole cell biocatalysis assay described in Example II. As shown in an HPLC chromatogram, one mutant ("R-922 UV mutant") showed a two to three fold increase in an ability to regioselectively oxidize avermectin to 4"-keto-avermectin as compared to wild-type strain R-922. Although the gene encoding the P450 monooxygenase responsible for the regioselectively oxidation activity, *ema1*, is not mutated in the R-922 UV mutant, this mutant nonetheless provides an excellent source for a cell-free extract containing ema1 protein.

EXAMPLE V

Isolation of the P450 Monooxygenase from *Streptomyces* Strain R-922

To enrich the P450 enzyme, 35 ml of active cell free extract were filtered through a 45 µm filter and fractionated by anion exchange chromatography. Anion exchange chromatography conditions were as follows:

FPLC instrument:	Äkta prime (from Pharmacia Biotech)
FPLC-column:	HiTrap™Q (5 ml) stacked onto Resource® Q (6 ml) (from Pharmacia Biotech)
eluents	buffer A: 25 mM Tris/HCl (pH 7.5)
	buffer B: 25 mM Tris/HCl (pH 7.5) containing 1 M KCl
temperature	eluent bottles and fractions in ice bath,
flow	3 ml/min
detection	UV 280nm
Pump table:	0.0 min 100% A 0% B
linear gradient to	2.0 min 90% A 10% B
	5.0 min 90% A 10% B

*linear gradient to*30.0 min	50% A	50% B
*linear gradient to*40.0 min	0% A	100% B
50.0min	0% A	100% B

Enzyme activity eluted with 35%-40% buffer B. The active fractions were pooled and concentrated by centrifugal filtration through Biomax™ filters with an exclusion limit of 5kD (commercially available from Millipore Corp., Bedford, MA) at 5000 rpm and then rediluted in disruption buffer containing 20% glycerol to a volume of 5 ml containing 3-10 mg/ml protein. This enriched enzyme solution contained at least 25% of the original enzyme activity.

The enzyme was further purified by size exclusion chromatography. Size exclusion chromatography conditions were as follows:

FPLC instrument:	Äkta prime (from Pharmacia Biotech)
FPLC-column:	HiLoad 26/60 Superdex® 200 prep grade (from Pharmacia Biotech)
sample:	3-5 ml enriched enzyme solution from the anion chromatography step
sample preparation:	filtered through 45 µm filter
eluent buffer:	PP-buffer (pH 7.0) + 0.1 M KCl
temperature:	4°C
flow:	2 ml/min
detection:	UV 280nm

Enzyme activity eluted between 205-235 ml eluent buffer. The active fractions were pooled, concentrated by centrifugal filtration through Biomax™ filters with an exclusion limit of 5 kD (from Millipore) at 5000 rpm, and rediluted in disruption buffer containing 20% glycerol to form a solution of 0.5-1 ml containing 2-5mg/ml protein. This enriched enzyme solution contained 10% of the original enzyme activity. This enzyme preparation, when checked for purity by SDS page, (*see*, generally, Laemmli, U.K., *Nature* **227**:680-685, 1970 and Current Protocols in Molecular Biology, *supra*) and stained with Coomassie blue, showed one dominant protein band with a molecular weight of 45-50 kD, according to reference proteins of known molecular weight.

EXAMPLE VI

Attempted Isolation of P450 Monooxygenase Genes

From *Streptomyces* Strains R-922 and I-1529

Based on results described above that suggested the enzyme from strain R-922 that is responsible for the regiospecific oxidation of avermectin to 4"-keto-avermectin is a P450 monooxygenase, a direct PCR-based approach to clone P450 monooxygenase genes from this strain was initiated (*see*, generally, Hyun *et al.*, *J. Microbiol. Biotechnol.* **8**(3):295-299, 1998). This approach is based on the fact that all P450 monooxygenase enzymes contain highly conserved oxygen-binding and heme-binding domains that are also conserved at the nucleotide level. PCR primers were designed to prime to these conserved domains and to amplify the DNA fragment from P450 genes using R-922 or I-1529 genomic DNA as a template. The PCR primers used are shown in Table 1.

Table 1

O_2-Binding Domain Primers (5' to 3')*	**Degeneracy**	**SEQ ID NOs**
I A G H E T T		43
ATC GCS GGS CAC GAG ACS AC	8	44
V A G H E T T		45
GTS GCS GGS CAC GAG ACS AC	16	46
L A G H E T T		47
CTS GCS GGS CAC GAG ACS AC	16	48
L L L I A G H E T		49
TS CTS CTS ATC GCS GGS CAC GAG AC[&]	32	50
Heme-Binding Domain Primers (3' to 5')*		
H Q C L G Q N L A		51
GTG GTC ACG GAS CCS TGC TTG GAS CG[&]	8	52
F G H G V H Q C		53
AAG CCS GTG CCS CAS GTG GTC ACG	8	54
F G F G V H Q C		55
AAG GCS AAG CCS CAS GTG GTC ACG	8	56
F G H G I H Q C		57
AAG CCS GTG CCS TAG GTG GTC ACG	4	58

F G H G V H F C		59
AAG CCS GTG CCS CAS GTG AAG ACG	8	60

* The amino acid sequence is shown on the top line and the corresponding nucleotide sequence is shown below on the second line; S=G or C.

[&] This primer was described by Hyun *et al.*, *supra*

PCR amplification using any of the primers specific to nucleotide sequences encoding the O_2-binding domain with any of the primers specific to the nucleotide sequences encoding the heme-binding domain and genomic DNA from *Streptomyces* strains R-922 or I-1529 resulted in the amplification of an approximately 350 bp DNA fragment. This is exactly the size that would be expected from this PCR amplification due to the approximately 350 bp separation in P450 genes of the gene segments encoding the O_2-binding and heme-binding sites.

The 350 bp PCR fragments were cloned into the pCR2.1-TOPO TA cloning plasmid (commercially available Invitrogen, Carlsbad, CA) and transformed into *E. coli* strain TOP10 (Invitrogen, Carlsbad, CA). Approximately 150 individual clones from strains R-922 and I-1529 were sequenced to determine how many unique P450 gene fragments were represented. Analysis of the sequences revealed that they included 8 unique P450 gene fragments from strain R-922 and 7 unique fragments from I-1529.

Blast analysis (alignment of the deduced amino acid sequences of P450 gene-specific PCR fragments derived from *Streptomyces tubercidicus* strain R-922 and *Streptomyces* strain I-1529, respectively, and the P450 monooxygenase from *S. thermotolerans* that is involved in the synthesis of carbomycin (Stol-ORFA) (GenBank Accession No. D30759) by the program Pretty from the University of Wisconsin Package version 10.1 (Altschul *et al.*, *Nucl. Acids Res.* **25**:3389-3402). demonstrated that all of the unique P450 gene fragments from both the R-922 and I-1529 strains were derived from P450 genes and encoded the region between the O_2-binding and heme-binding domains.

Next, in order to clone the full-length genes from which the PCR fragments were derived, the DNA fragments cloned by PCR were used as hybridization probes to gene libraries containing genomic DNA from strains R-922 and I-1529. To do this, genomic DNA from the R-922 and I-1529 strains was partially digested with Sau3A I, dephosphorylated with

calf intestinal alkaline phosphatase (CIP) and ligated into the cosmid pPEH215, a modified version of SuperCos 1 (commercially available from Stratagene, La Jolla, CA). Ligation products were packaged using the Gigapack III XL packaging extract and transfected into *E. coli* XL1 Blue MR host cells. Twelve cosmids that strongly hybridized to the PCR-generated P450 gene fragments were identified from the R-922 library, from which three unique P-450 genes were subcloned and sequenced. The hybridizations were performed at high stringency conditions according to the protocol of Church and Gilbert (Church and Gilbert, *Proc. Natl. Acad. Sci. USA* **81**:1991-1995, 1984). In brief, these high stringency conditions include Hybrid Buffer containing 500 mM Na-phosphate, 1 mM EDTA, 7% SDS, 1% BSA; Wash Buffer 1 containing 40 mM Na-phosphate, 1 mM EDTA, 5% SDS, 0.5% BSA; and Wash Buffer 2 containing 40 mM Na-phosphate, 1 mM EDTA, 1% SDS (Note that other high stringency hybridizations conditions are described, for example, in Current Protocols in Molecular Biology, *supra.*) Nineteen strongly hybridizing cosmids were identified from the I-1529 library, and from these, four unique P-450 genes were subcloned and sequenced.

In yet a further approach to isolate diverse P450 monooxygenase genes from strains R-922 and I-1529, a known P450 gene from another bacterium was used as a hybridization probe to identify cosmid clones containing homologous P450 genes from strains R-922 and I-1529. The *epoF* P450 gene from *Sorangium cellulosum* strain So ce90 that is involved in the synthesis of epothilones (Molnar *et al.*, *Chem Biol.* 7(2):97-109, 2000) was used as a probe in this effort. Using the *epoF* P450 gene probe, one cosmid was identified from strain R-922 (clone LC), and threewere identified from strain I-1529 (clones LA, LB, and EA). In each case, the homologous gene fragment was subcloned and sequenced, and found to code for P450 monooxygenase enzymes.

However, a comparison of the 17 polypeptide sequences identified in Example VII (below) failed to match any of these cloned genes. Two of the polypeptide sequences (namely, LVKDDPALLPR and AVHELMR) mapped to the region between the O_2 and heme binding domains, and so these should have identified any of the partial gene fragments derived by the PCR approach. Thus, the standard approaches based on the known PCR technique of Hyun et al., *supra*, and using known P450 genes as hybridization probes failed to identify the gene that encodes the specific P450 monooxygenase responsible for the regioselective

oxidation of avermectin. Accordingly, it was determined that additional experimentation was required to isolate the gene encoding the P450 monooxygenase of the invention.

EXAMPLE VII

Partial Sequencing of the P450 Monooxygenase from *Streptomyces* Strain R-922

Partial amino acid sequencing of the P450 monooxygenase from *Streptomyces* strain R-922 was carried out by the Friedrich Miescher Institute, Basel Switzerland. The protein of the dominant band on the SDS page was tryptically digested and the formed peptides separated and sequenced by mass spectrometry and Edman degradation (*see*, generally, Zerbe-Burkhardt *et al.*, *J. Biol. Chem.* **273**:6508, 1998). The sequence of the following 17 peptides were found:

Sequence	Sequence I.D. No.
HPGEPNVMDPALITDPFTGYGALR	(SEQ ID NO:61)
FVNNPASPSLNYAPEDNPLTR	(SEQ ID NO:62)
LLTHYPDISLGIAPEHLER	(SEQ ID NO:63)
VYLLGSILNYDAPDHTR	(SEQ ID NO:64)
TWGADLISMDPDR	(SEQ ID NO:65)
EALTDDLLSELIR	(SEQ ID NO:66)
FMDDSPVWLVTR	(SEQ ID NO:67)
LMEMLGLPEHLR	(SEQ ID NO:68)
VEQIADALLAR	(SEQ ID NO:69)
LVKDDPALLPR	(SEQ ID NO:70)
DDPALLPR	(SEQ ID NO:71)
TPLPGNWR	(SEQ ID NO:72)
LNSLPVR	(SEQ ID NO:73)
ITDLRPR	(SEQ ID NO:74)
EQGPVVR	(SEQ ID NO:75)
AVHELMR	(SEQ ID NO:76)

AFTAR (SEQ ID NO:77)

FEEVR (SEQ ID NO:78)

Alignment of these peptides to a selection of actinomycete P450 monooxygenase sequences indicated that all the peptides were fragments of a single P450 mono-oxygenase.

EXAMPLE VIII

Cloning the P450 Monooxygenase Gene from Strain R-922 that Encodes the Enzyme Responsible for the Oxidation of Avermectin to 4"-Keto-Avermectin

PCR primers were designed by reverse translation from the amino acid sequences of several of the peptides derived from the P450 enzyme of strain R-922 (*see* Example VII and Table 2 below). Each of five forward primers (2aF, 2bF, 3F, 1F, and 7F) was paired with one reverse primer (5R) in PCR reactions with R-922 genomic DNA as a template. In each reaction, a DNA fragment of the expected size was produced.

Table 2

Primer	Primer sequence and the amino acid sequence to which they were designed*	Degen-eracy	Expected size (bp)**	SEQ ID NO:
2aF	P G E D N V M 5'-CCS GGS GAR CCS AAY GTS ATG-3'	64	600	79 80
2bF	A L I T D P F 5'-GCS CTS ATY ACS GAC CCS TTC-3'	32	580	81 82
3F	F M D D S P V W 5'-TTC ATG GAC GAC WSS CCS GTS TGG-3'	32	549	83 84
1F	L N Y D A P D H 5'-CTS AAY TAY GAC GCS CCS GAC CAC-3'	32	350	85 86
7F	V E Q I A D A L 5'-GTS GAR CAG ATY GCS GAC GCS CTS-3'	32	300	87 88
5R	D L I S M D P D 3'-CTG GAS TAR WSS TAC CTG GGS CTG-5'	64	---	89 90

* Ambiguity codes: Y=C or T; R=A or G; S=C or G; W=A or T

** Expected size of PCR product when the primer is when paired with primer 5R

The 580 and 600 bp PCR fragments generated by using primers (2bF and 5R) and (2aF and 5R), respectively, were cloned into the pCR-Blunt II-TOPO cloning plasmid (commercially available from Invitrogen, Carlsbad, CA) and transformed into *E. coli* strain TOP10 (Invitrogen, Carlsbad, CA). The inserted DNA fragments were then sequenced. Examination of the sequences revealed that the 600 and 580 bp fragments were identical in the 580 bp of sequence that they have in common. Also, there was a perfect match between the deduced amino acid sequence (SEQ ID NO:2) derived from the nucleotide sequence of the 600 bp and 580 bp fragments and the amino acid sequences of peptides isolated from the purified $P450_{Ema1}$ enzyme that aligned in this region of the isolated gene. This result strongly suggested that the gene fragments isolated in these clones are derived from the gene that encodes the $P450_{Ema1}$ enzyme that is responsible for the oxidation of avermectin to 4"-keto-avermectin.

The 600 bp PCR fragment produced using primers 2aF (SEQ ID No:80) and 5R (SEQ ID No:90) was used as a hybridization probe to a cosmid library of genomic DNA isolated from strain R-922 (cosmid library described in Example VI). Two cosmids named pPEH249 and pPEH250 were identified that hybridized strongly with the probe. The portion of each cosmid encoding the P450 enzyme was sequenced and the sequences were found to be identical between the two cosmids. The complete coding sequence of the *ema1* gene was identified (SEQ ID NO:1). The amino acid sequence of all polypeptide fragments from $P450_{Ema1}$ matched perfectly with the deduced amino acid sequence from the *ema1* gene. Comparison of the deduced amino acid sequence of the protein encoded by the *ema1* gene using BLASTP (Altschul *et al.*, *supra*) determined that the closest match in the databases is to a P450 monooxygenase from *S. thermotolerans* that has a role in the biosynthesis of carbomycin (Arisawa *et al.*, *Biosci. Biotech. Biochem.* **59**(4):582-588, 1995) and whose identity with *ema1* is only 49% (Identities = 202/409 (49%), Positives = 271/409 (65%), Gaps = 2/409 (0%)). In the Blast analysis, the following settings were employed:

```
BLASTP 2.0.10
Lambda      K       H
   0.322    0.140    0.428
Gapped
Lambda      K       H
   0.270   0.0470    0.230
Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 375001765
Number of Sequences: 1271323
Number of extensions: 16451653
Number of successful extensions: 46738
Number of sequences better than 10.0: 2211
Number of HSP's better than 10.0 without gapping: 628
Number of HSP's successfully gapped in prelim test: 1583
Number of HSP's that attempted gapping in prelim test: 43251
Number of HSP's gapped (non-prelim): 2577
length of query: 430
length of database: 409,691,007
effective HSP length: 55
effective length of query: 375
effective length of database: 339,768,242
effective search space: 127413090750
effective search space used: 127413090750
```

A similar comparison of the nucleotide sequences of these two genes demonstrated that they are 65% identical at the nucleotide level. These results demonstrate that $P450_{Ema1}$ is a new enzyme.

EXAMPLE IX

Heterologous Expression of the *ema1* Gene in *Streptomyces lividans* Strain ZX7

The coding sequence of the *ema1* gene was fused to the thiostrepton-inducible promoter (*tipA*) (Murakami *et al.*, *J. Bacteriol.* **171**:1459-1466, 1989). The *tipA* promoter was derived from plasmid pSIT151 (Herron and Evans, *FEMS Microbiology Letters* **171**:215-221, 1999).

The fusion of the *tipA* promoter and the *ema1* coding sequence was achieved by first amplifying the *ema1* coding sequence with the following primers to introduce a PacI cloning site at the 5' end and a PmeI compatible end on the 3' end.

Forward Primer: The underlined sequence is a PacI recognition sequence; the sequence in bold-face type is the start of the coding sequence of *ema1*.

5'-AGATTAATTAA**TGTCGGAATTAATGAACTGTCCGTT**-3' (SEQ ID NO:91)

Reverse Primer: The underlined sequence is half of a PmeI recognition sequence; the bold-face type sequence is the reverse complement of the *emal* translation stop codon followed by the 3' end of the *emal* coding sequence.

5'-AAAC**TCA**CCCCAACCGCACCGGCAGCGAGTTC-3" (SEQ ID NO:92)

The PacI-digested PCR fragment containing the *emal* coding sequence was cloned into plasmid pTBBKA (*see* Figure 1) that was restricted (*i.e.*, digested) with PacI and PmeI, and the ligated plasmid transformed into *E. coli*. Four clones were sequenced. Three of the four contained the complete and correct *emal* coding sequence. The fourth *emal* gene clone contained a truncated version of the *emal* gene. The full-length *emal* gene encodes a protein that begins with the amino acid sequence MSELMNS (SEQ ID NO:93). The truncated gene encodes a protein that lacks the first 4 amino acids and begins with the second methionine residue. This gene has been named *emalA*. The nucleotide and amino acid sequence of *emalA* are provided as SEQ ID NO:33 and SEQ ID NO:34, respectively. The *emal* and *emalA* genes in these plasmids, pTBBKA-*emal* and pTBBKA-*emalA*, are in the correct juxtaposition with the *tipA* promoter to cause expression of the genes from this promoter.

Plasmid pTBBKA contains a gene from the *Streptomyces* insertion element IS117 that encodes an integrase that catalyzes site-specific integration of the plasmid into the chromosome of *Streptomyces* species (Henderson *et al.*, *Mol. Microbiol.* **3:**1307-1318, 1989 and Lydiate *et al.*, *Mol. Gen. Genet.* **203:**79-88, 1986). Since plasmid pTBBKA has only an *E. coli* replication origin and contains a mobilization site, it can be transferred from *E. coli* to *Streptomyces* strains by conjugation where it will not replicate. However, it is able to integrate into the chromosome due to the IS117 integrase and *Streptomyces* clones containing chromosomal integrations can be selected by resistance to kanamycin due to the plasmid-borne kanamycin resistance gene.

The *emal* coding sequence was also cloned into other plasmids that are either replicative in *Streptomyces* or, like pTBBKA, integrate into the chromosome upon introduction into a *Streptomyces* host. For example, *emal* was cloned into plasmid pEAA, which is similar to plasmid pTBBKA but the KpnI/PacI fragment containing the *tipA* promoter was replaced with the *ermE* gene promoter (Schmitt-John and Engels, *Appl*

Microbiol Biotechnol. **36**(4):493-498, 1992). In addition, pEAA does not contain the kanamycin resistance gene. The *ema1* gene was cloned into pEAA as a PacI/PmeI fragment to create plasmid pEAA-*ema1* in which the *ema1* gene is expressed from the constitutive *ermE* promoter.

Plasmid pTUA1A is a *Streptomyces-E.coli* shuttle plasmid (*see* Figure 2) that contains the *tipA* promoter. The *ema1* gene was also cloned into the PacI/PmeI site in plasmid pTUA1A to create plasmid pTUA-*ema1*.

The *ema1A* gene fragment was also ligated as a PacI/PmeI fragment into plasmids pTUA1A, and pEAA in the same way as the *ema1* gene fragment to create plasmids pTUA-ema1A, and pEAA-ema1A, respectively.

The pTBBKA, pTUA1A, and pEAA based plasmids containing the *ema1* or *ema1A* genes were introduced into *S. lividans* ZX7 and in each case transformants were obtained and verified (*S. lividans* strains ZX7::pTBBKA-*ema1* or *ema1A*, ZX7 (pTUA-*ema1* or *–ema1A)*, and ZX7::pEAA-*ema1* or *-ema1A*, respectively).

Wild-type *Streptomyces lividans* strain ZX7 was tested and found to be incapable of the oxidation of avermectin to 4"-keto-avermectin. Transformed *S. lividans* strains ZX7::pTBBKA-*ema1*, ZX7::pTBBKA-*ema1A*, ZX7 (pTUA-*ema1*), ZX7 (pTUA-*ema1A*), ZX7::pEAA-*ema1*, and ZX7::pEAA-*ema1A* were each tested for the ability to oxidize avermectin to 4"-keto-avermectin using resting cells. To do this, the whole cell biocatalysis assay described above (including analysis method) was performed. Note that for the whole cell biocatalysis assay, transformed *Streptomyces lividans*, like strain R-922, was grown in PHG medium and, again like strain R-922, had a reaction time of 16 hours (*i.e.*, during which time the 500 mg transformed *Streptomyces lividans* wet cells in 10 ml of 50 mM potassium phosphate buffer, pH 7.0, were shaken at 160 rpm at 28°C in the presence of 15 µl of a solution of avermectin in isopropanol (30 mg/ml)).

In the presence of the inducer, thiostrepton (5 ug/ml), the *ema1*- or *ema1A*-containing strains ZX7::pTBBKA-*ema1*, ZX7::pTBBKA-*ema1A*, ZX7 (pTUA-*ema1*), ZX7 (pTUA-*ema1A*) were found to oxidize avermectin to 4"-keto-avermectin as evidenced by the appearance of the oxidized 4"-keto-avermectin compound (*see* Table 3).

Table 3

Beispiel 1: Strain	% Conversion of Avermectin	
	2 hour	**16 hour**
***Streptomyces lividans* ZX7 + Plasmid**[1]		
None	0	0
pTBBKA-*ema1A*	0.5 ±0.059	1.17 ±0.112
pTBBKA-*ema1*	0.21 ±0.0.356	0.65 ±0.079
pTUA-*ema1*	20.96 ±1.044	42.0 ±2.5
pEAA-*ema1*	3.0 ±0.232	24.1 ±0.358
pTBBKA-*ema2*	4.79 ±0.096	9.57 ±0.423
pTUA-*ema2*	0.77 ±0.138	2.05 ±0.537
pEAA-*ema2*	0.0	1.73 ±3.00
pTBBKA-*ema1/fd233*	8.89 ±0.720	30.99 ±0.880
pTUA-*ema1/fd233*	23.29 ±0.854	61.2 ±3.548
pEAA-*ema1/fd233*	8.26 ±0.845	10.66 ±0.858
pTUA-*ema2/fd233*	1.85 ±0.861	6.40 ±1.918
***Pseudomonas putida* S12 + Plasmid**		
None		0
pRK-*ema1*	ND[2]	18
pRK-*ema1/fd233*	ND	32

[1]pTBBKA= IS117 integrase, tipA promoter; pTUA= replicative plasmid, tipA promoter; pEAA= IS117 integrase, ermE promoter

[2]Not Determined

These results conclusively demonstrate that the $P450_{Ema1}$ enzyme encoded by the *ema1* gene is responsible for the oxidation of avermectin to 4"-keto-avermectin in *S. tubercidicus* strain R-922. Furthermore, the data demonstrates that the *ema1A* gene that is 4 amino acids shorter on the N-terminus than the native *ema1* gene also encodes an active $P450_{Ema1}$ enzyme. As can be demonstrated by HPLC analysis, oxidation of avermectin to 4"-keto-avermectin by *S.*

lividans strain ZX7::pTBBKA-*ema1* following induction of ema1 expression with 0, 0,5, or 5.0 μg/ml thiostrepton. is variable depending upon the amount of thiostrepton used to induce expression of *ema1*. Note that *S. lividans* strains ZX7::pEAA-*ema1* and ZX7::pEAA-*ema1A* (*see* Table 3) demonstrated this oxidation activity in the absence of thiostrepton since in these strains the *ema1* or *ema1A* genes are expressed from the *ermE* promoter that does not require induction.

EXAMPLE X

Isolation of an *ema1*-Homologous Gene From *Streptomyces tubercidicus* Strain I-1529

Streptomyces tubercidicus strain I-1529 was also found to be active in biocatalysis of avermectin to form the 4"-keto-avermectin derivative. The cosmid library from strain I-1529, described in Example VI, was probed at the high stringency conditions of Church and Gilbert (Church and Gilbert, *Proc. Natl. Acad. Sci. USA* **81**:1991-1995, 1984) with the 600 bp *ema1* PCR fragment produced using primers 2aF (SEQ ID No:80) and 5R (SEQ ID No:90) described previously to identify clones containing the *ema1* homolog from strain I-1529. Three strongly hybridizing cosmids were identified. The P450 gene regions in two of the cosmids, pPEH252 and pPEH253, were sequenced and found to be identical. Analysis of the DNA sequence revealed the presence of a gene with high homology to the *ema1* gene of strain R-922. A comparison of the deduced amino acid sequence of Ema2 (*i.e.*, $P450_{Ema2}$), Ema1 (*i.e.*, $P450_{Ema1}$), and a P450 monooxygenase from *Streptomyces thermotolerans* that is involved in the biosynthesis of carbomycin (Carb-450) (GenBank Accession No. D30759). demonstrated that all of the unique P450 gene fragments from both the R-922 and I-1529 strains were derived from P450 genes and encoded the region between the O_2-binding and heme-binding domains.

The gene from *Streptomyces tubercidicus* strain I-1529, named *ema2*, encodes an enzyme with 90% identity at the amino acid level and 90.6% identity at the nucleotide level to the $P450_{Ema1}$ enzyme. The nucleotide sequence of the *ema2* gene and the deduced amino acid sequence of $P450_{Ema2}$ are provided in SEQ ID NO:3 and SEQ ID NO:4, respectively.

The *ema2* coding sequence was cloned in the same manner as the *ema1* and *ema1A* genes into plasmids pTBBKA, pTUA1A, and pEAA such that the coding sequence was

functionally fused to the *tipA* or *ermE** promoter in these plasmids. The resulting plasmids, pTBBKA-*ema2*, pTUA-*ema2*, and pEAA-*ema2* were transferred from *E. coli* to *S. lividans* ZX7 by conjugation to create strains ZX7::TBBKA-*ema2* and ZX7 (pTUA-*ema2*), and ZX7::pEAA-*ema2* containing the *ema2* gene integrated into the chromosome or maintained on a plasmid.

Strains ZX7::TBBKA-*ema2*, ZX7 (pTUA-*ema2*), and ZX7::pEAA-*ema2* were next tested for the ability to oxidize avermectin to 4"-keto-avermectin. The *ema2* gene was also shown to provide biocatalysis activity, although at a lower level compared to the *ema1* gene (*see* Table 3).

These results demonstrate that the *ema2* gene from *S. tubercidicus* strain I-1529 also encodes a P450 enzyme ($P450_{Ema2}$) capable of oxidizing avermectin to 4"-keto-avermectin.

EXAMPLE XI

Characterization of ema1 Homologs From Other Biocatalysis Strains

Seventeen *Streptomyces* sp. strains, including strains R-922 and I-1529, were identified that are capable of catalyzing the regiospecific oxidation of the 4"-carbinol of avermectin to a ketone. Next, the isolation and characterization of the genes encoding the biocatalysis enzyme from all of these strains was accomplished.

To do this, genomic DNA was isolated from the strains and was evaluated by restriction with several restriction endonucleases and Southern hybridization with the *ema1* gene. A specific restriction endonuclease was identified for each DNA that would generate a single DNA fragment of a defined size to which the *ema1* gene hybridizes. For each strain, there was only one strongly hybridizing DNA fragment, thus suggesting that other P450 genes were not detected under the high stringency hybridization conditions used in these experiments. Each DNA was digested with the appropriate restriction endonuclease, and the DNA was subjected to agarose gel electrophoresis. DNA in a narrow size range that included the size of the *ema1*-hybridizing fragment was excised from the gel. The size selected DNA was ligated into an appropriate cloning plasmid and this ligated plasmid was used to transform *E. coli*. The *E. coli* clones from each experiment were screened by colony hybridization with the *ema1* gene fragment to identify clones containing the *ema1*-homologous DNA fragment.

The nucleotide sequence of the cloned DNA in each *ema1*-homologous clone was determined and examined for the presence of a gene encoding a P450 enzyme with homology to *ema1*. In this way, *ema1*-homologous genes were isolated from 14 of the 15 other active strains. The nucleotide and deduced amino acid sequences of these are referenced in Table 4 as SEQ ID NOS:5-32 and 94-95. The relationship of these enzymes can be shown in the form of a phylogenetic tree. Such a phylogenetic tree can be generated using the commercially available GCG Wisconsin software program version 1.0 (Madison, WI).

Table 4

Strain Number		Classification	SEQ ID NO (nucleotide and amino acid, respectively)
R-0922	*ema1*	*Streptomyces tubercidicus*	2. 1 and 2
I-1529	*ema2*	*Streptomyces tubercidicus*	3 and 4
1053	*ema3*	*Streptomyces rimosus*	5 and 6
R-0401	*ema4*	*Streptomyces lydicus*	7 and 8
I-1525	*ema5*	*Streptomyces sp.*	9 and 10
DSM-40241	*ema6*	*Streptomyces chattanoogensis**	3. 11 and 12
IHS-0435	*ema7*	*Streptomyces sp.*	13 and 14
C-00083	*ema8*	*Streptomyces albofaciens*	15 and 16
MAAG-7479	*ema9*	*Streptomyces platensis*	17 and 18
A/96-1208710	*ema10*	*Streptomyces kasugaensis*	4. 19 and 20
R-2374	*ema11*	*Streptomyces rimosus*	21 and 22
MAAG-7027	*ema12*	*Streptomyces tubercidicus*	5. 23 and 24
Tue-3077	*ema13*	*Streptomyces platensis*	25 and 26
I-1548	*ema14*	*Streptomyces platensis*	27 and 28
NRRL-2433	*ema15*	*Streptomyces lydicus*	6. 29 and 30
MAAG-0114	*ema16*	*Streptomyces lydicus*	31 and 32
DSM-40261	*ema17*	*Streptomyces tubercidicus*	94 and 95

* This strain was shown to be in the *chattanoogensis* species by 16s rDNA analysis; however, classical taxonomic methods used by the German culture collection (DSMZ) showed it to be *saraceticus*.

EXAMPLE XII

Construction of His-tagged *ema1* and *ema1* Homologs to Facilitate Enzyme Purification

In order to purify the $P450_{Ema1}$ enzyme and the P450 enzymes encoded by the ema1 homologs from other biocatalysis strains, each of the P450 genes was cloned into the *E. coli* expression plasmid pET-28b(+) (commercially available from Novagen, Madison, WI). The pET-28 plasmids are designed to facilitate His-tag fusions at either the N-, or C-terminus and to provide strong expression of the genes in *E. coli* from the T7 phage promoter. In many cases, the coding sequence of the *ema* genes begins with the sequence ATGT. These genes were amplified by PCR such that the primers on the 5' end incorporated a PciI recognition site (5' ATATGT 3') at the 5' terminus. The last four bases of the PciI site correspond to the ATGT at the beginning of the *ema* gene coding sequence.

PCR primers at the 3' end of the genes were designed to remove the translation stop codon at the end of the *ema* gene coding sequence and to add an XhoI recognition site to the 3' terminus. The resulting PCR fragments were restricted with PciI and XhoI to generate PciI ends at the 5' termini and XhoI ends at the 3' termini, thereby facilitating cloning of the fragments into pET-28b(+) previously restricted with NcoI and XhoI. Since PciI and NcoI ends are compatible, the fragments were cloned into pET-28b(+) in the proper orientation to the T7 promoter and ribosome binding site in the plasmid to provide expression of the genes.

At the 3' end of each *ema* gene, the coding sequence was fused in frame at the XhoI site to the His-tag sequence followed by a translation stop codon. This results in the production of an Ema enzyme with six histidine residues added to the C-terminus to facilitate purification on nickel columns.

In the case of *ema* genes in which the ATG translation initiation codon is not followed by a T nucleotide, the *ema* genes were amplified by PCR using a different strategy for the 5' end. The primers at the 5' end were designed to incorporate a C immediately preceding the ATG translation initiation codon and the primers at the 3' end were the same as described above. The PCR fragments that were amplified were restricted with XhoI to create an XhoI end at the 3'-terminus and the 5' end was left as a blunt end. These fragments were cloned into pET-28b(+) that had been restricted with NcoI, but the NcoI ends were made blunt-ended by treatment with mung bean exonuclease, and restricted with XhoI.

In this manner, the *ema* genes were cloned into pET-28b(+) to create a functional fusion with the T7 promoter and the His-tag at the C-terminus as described previously. All His-tagged *ema* genes were sequenced to ensure that no errors were introduced by PCR.

Large amounts of the $P450_{Ema1}$ and $P450_{Ema2}$ enzymes were isolated and purified by standard protocols. *E. coli* strain BL21 DE3 (commercially available from Invitrogen; Carlsbad, CA) containing the T7 RNA polymerase gene under the control of the inducible *tac* promoter and the appropriate pET-28/*ema* plasmid was cultured and the cells were harvested and lysed. The lysates were applied to Ni-NTA columns (commercially available from Qiagen Inc., Valencia, CA) and the protein were purified according to the procedure recommended by the manufacturer.

Purified His-tagged $P450_{Ema1}$ and $P450_{Ema2}$ were highly active in *in vitro* activity assays as evidenced by a high rate of conversion of avermectin to 4"-keto-avemectin.

EXAMPLE XIII

Expression of *ema1* in *Pseudomonas*

The ema1 gene constructs were next introduced into *P. putida* (wildtype *P. putida* commercially available from the American Type Culture Collection, Manassas, Virginia; ATCC Nos. 700801 and 17453). The *ema1* and *ema1/fd233* gene fragments were cloned as PacI/PmeI fragments into the plasmid pUK21 (Viera and Messing, *Gene* **100**:189-194, 1991). The fragments were cloned into a position located between the tac promoter (P_{tac}) and terminator (T_{tac}) on pUK21 in the proper orientation for expression from the tac promoter. The P_{tac}-*ema1*-T_{tac} and P_{tac}-*ema1/fd233*-T_{tac} gene fragments were removed from pUK21 as BglII fragments and these were cloned into the broad host-range, transmissible plasmid, pRK290 (Ditta *et al.*, *Proc. Natl. Acad. Sci. USA* **77**:7347-7351, 1980) to create plasmids pRK-*ema1* and pRK-*ema1/fd233* (Figure 3). These plasmids were introduced into *P. putida* strains ATCC 700801 and ATCC 17453 by conjugal transfer from *E. coli* hosts by standard methodology (Ditta *et al.*, *Proc. Natl. Acad. Sci. USA* **77**:7347-7351, 1980).

P. putida ATCC 700801 and ATCC 17453 containing plasmids pRK-*ema1* or pRK-*ema1/fd233* were tested for the ability to catalyze the oxidation of avermectin. The results shown in Table 3 demonstrate that these strains are able to catalyze this reaction.

EXAMPLE XIV

Identification of Genes Encoding Ferredoxins That Are Active With the $P450_{Ema1}$ Monooxygenase

P450 monooxygenases require two electrons for each hydroxylation reaction catalyzed (Mueller *et al.*, "Twenty-five years of $P450_{cam}$ research: Mechanistic Insights into Oxygenase Catalysis." Cytochrome P450, 2nd Edition, P.R. Ortiz de Montellano (ed.), pp. 83-124; Plenum Press, NY 1995). These electrons are transferred to the P450 monooxygenase one at a time by a ferredoxin. The electrons are ultimately derived from NAD(P)H and are passed to the ferredoxin by a ferredoxin reductase. Specific P-450 monooxygenase enzymes have a higher activity when they interact with a specific ferredoxin. In many cases, the gene encoding a ferredoxin that interacts specifically with a given P450 monooxygenase is located adjacent to the gene encoding the P450 enzyme.

As described above, in addition to the *ema1* gene, four P450 genes from strain R-922 and seven P450 genes from strain I-1529 (*see* Example VI) were isolated and sequenced. In some of these, there was sufficient sequence information about the DNA flanking the P-450 genes to look for the presence of associated ferredoxin genes. By this approach, two unique ferredoxin genes were identified from each of the two strains. Ferredoxin genes *fd229* and *fd230* were identified from strain R-922, and *fd233* and *fd*EA were identified from strain I-1529. In addition, a ferredoxin reductase gene was found to reside adjacent to the *fd*EA gene from strain I-1529.

In order to test the biological activity of each of these ferredoxins in combination with $P450_{Ema1}$, each individual ferredoxin gene was amplified by PCR to produce a gene fragment that included a blunt 5'-end, the native ribosome-binding site and ferredoxin gene coding sequence, and a PmeI restriction site on the 3'-end. Each such ferredoxin gene fragment was cloned into the PmeI site located 3' to the *ema1* gene in plasmid pTUA-*ema1*. In this way, artificial operons consisting of the *ema1* gene and one of the ferredoxin genes operably linked to a functional promoter were created.

In the case of the *fd*EA ferredoxin gene in which a ferredoxin reductase gene, *fre*EA, was found to be located adjacent to the *fd*EA gene, a DNA fragment containing both the *fd*EA and *fre*EA genes was generated by a similar PCR strategy. This gene fragment was also cloned in the PmeI site of plasmid pTUA-*ema1* as described for the other ferredoxin genes.

Each *ema1*-ferredoxin gene combination was tested for biological activity by introduction of the individual *ema1*-ferredoxin gene plasmids into *S. lividans* strain ZX7. The biocatalysis activity derived from each plasmid in *S. lividans* was determined. Of the four different constructs, only the ferredoxin gene *fd233* derived from strain I-1529 provided increased activity when compared to the expression of *ema1* alone in the same plasmid and host background (*see* Table 3). The pTUA-*ema1/fd233* plasmid in *S. lividans* provided approximately 1.5 to 3- fold higher activity compared to the pTUA-*ema1* plasmid. The other three plasmids containing the other ferredoxin genes gave results essentially the same as the plasmid with only the *ema1* gene. Likewise, the pTUA-*ema1/fd*EA/*fre*EA plasmid did not yield results different from those of pTUA-*ema1*. The nucleotide and deduced amino acid sequences of the *fd233* gene are shown in SEQ ID NOs:35 and 36, respectively.

A BLAST analysis of the nucelotide and amino acid sequences of *fd233* revealed that the closest matches were to ferredoxins from *S. coelicolor* (GenBank Accession AL445945) and *S. lividans* (GenBank Accession AF072709). At the nucleotide level, *fd233* shares 80 and 79.8 % identity with the ferredoxin genes from *S. coelicolor* and *S. lividans*, respectively. At the peptide level, *fd233* shares 79.4 and 77.8% identity with the ferredoxins from *S. coelicolor* and *S. lividans*, respectively.

Since *fd233* is derived from strain I-1529 and *ema1* is from strain R-922, the proteins encoded by the two genes cannot interact with each other in nature. In an approach designed to identify a ferredoxin gene from strain R-922 that is homologous to the *fd233* gene and that might encode a ferredoxin that interacts optimally with the $P450_{Ema1}$, the *fd233* gene was used as a hybridization probe to a gene library of DNA from strain R-922. A strongly hybridizing cosmid, pPEH232, was identified and the hybridizing DNA was cloned and sequenced. Comparison of the deduced amino acid sequences from *fd233* and the ferredoxin gene on cosmid pPEH232, *fd232*, revealed that they differed in only a single amino acid.

In a similar manner, plasmid pTUA-*ema1-fd232* was constructed and tested in *S. lividans* ZX7. This plasmid gave similar results as those obtained with plasmid pTUA-*ema1-fd233* (*see* Table 3). The nucleotide and deduced amino acid sequences of *fd232* are shown in SEQ ID NOs:37 and 38, respectively.

The *ema1-fd233* operon was also subcloned, as a PacI-PmeI fragment, into pTBBKA and pEAA that had been digested with the same restriction enzymes. *S. lividans*

ZX7::pTBBKA-*ema1-fd233*, and S. *lividans* ZX7::pEAA-*ema1-fd233* were tested in the avermectin conversion assay and found to have higher activities than the strains harboring the *ema1* gene alone in the comparable plasmids (*see* Table 3).

EXAMPLE XV

Heterologous Expression of $P450_{Ema1}$ and $P450_{Ema2}$ in Other Cells

The expression constructs pRK-*ema1* (Example XIII) and pRK-*ema2* (created in a way analogous to that described in Example XIII for pRK-*ema1*) were mobilized by conjugation into three fluorescent soil *Pseudomonas* strains. Conjugation was performed according to standard methods (Ditta *et al.*, *Proc. Natl. Acad. Sci. USA* **77**:7347-7351, 1980). The strains were: *P. fluorescens* MOCG134, *P. fluorescens* Pf-5, and *P. fluorescens* CHAO. Standard resting cell assays for the conversion of avermectin to 4"-ketoavermectin were conducted for each of the transconjugants. For strains Pf-5 and CHAO, the levels of conversion were below the detection limit. Strain MOCG134 yielded 3% conversion for *ema1* and 5% for *ema2*.

In addition, the constructs listed in the Table 5 were introduced into *Streptomyces avermitilis* MOS-0001 by protoplast-mediated transformation (Kieser, T.; Bibb, M.J.; Buttner, M.J.; Chater, K.F.; Hopwood, D.A. (eds.): Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich (England), 2000), (Stutzman-Engwall, K. *et al.* (1999) *Streptomyces avermitilis* gene directing the ratio of B_2:B_1 avermectins, WO 99/41389).

Table 5

Construct	**% Conversion of avermectin, 16 hrs**
None	0
pTBBKA-*ema1*	10.90 +/- 3.48
pTUA-*ema1*	5.326 +/- 2.19
pEAA-*ema1*	6.74 +/- 0.08
pTBBKA-*ema1A/fd233*	28.50 +/- 0.20
pTUA-*ema1A/fd233*	23.97 +/- 5.95

Wild-type *Str. avermitilis* MOS-0001 was tested and found to be incapable of the oxidation of avermectin to 4"-ketoavermectin.

Transformed *S. avermitilis* strains MOS-0001::pTBBKA-*ema1*, MOS-0001 (pTUA-*ema1*), MOS-0001::pEAA-*ema1*, MOS-0001::pTBBKA-*ema1A/fd233*, and MOS-0001 (pTUA-*ema1A/fd233*) were each tested for their ability to oxidize avermectin to 4"-keto-avermectin using resting cells. To do this, the whole cell biocatalysis assay described above (including analysis method) was performed. Note that for the whole cell biocatalysis assay, transformed *Streptomyces avermitilis*, like strain R-922, was grown in PHG medium and, again like strain R-922, had a reaction time of 16 hours (*i.e.*, during which time the 500 mg transformed *Streptomyces avermitilis* wet cells in 10 ml of 50 mM potassium phosphate buffer, pH 7.0, were shaken at 160 rpm at 28°C in the presence of 15 µl of a solution of avermectin in isopropanol (30 mg/ml)).

As shown in Table 5, in the presence of the inducer, thiostrepton (5 µg/ml), the *ema1*- or *ema1A/fd233*-containing strains MOS-0001::pTBBKA-*ema1*, MOS-0001::pTBBKA-*ema1A/fd233*, MOS-0001 (pTUA-*ema1*), MOS-0001 (pTUA-*ema1A/fd233*) were found to oxidize avermectin to 4"-keto-avermectin as evidenced by the appearance of the oxidized 4"-keto-avermectin compound. Note that the *S. avermitilis* strain MOS-0001::pEAA-*ema1* demonstrated this oxidation activity in the absence of thiostrepton since in this strain the *ema1* gene is expressed from the *ermE* promoter that does not require induction.

Thus, expression of the *ema1* P450 monooxygenase gene in various *Streptomyces* and *Pseudomonas* strains provided recombinant cells that were able to convert avermectin to 4"-ketoavermectin in resting cell assays.

Next, expression and activity of $P450_{Ema1}$ monooxygenase was tested in *E. coli*. To do this, the *ema1* gene was cloned into the *E. coli* expression plasmid pET-28b(+) (commercially available from Novagen, Madison, WI) as described previously. *E. coli* strain BL21 DE3 (commercially available from Invitrogen; Carlsbad, CA) that contains the T7 RNA polymerase gene under control of the inducible *tac* promoter and the pET-28/*ema1* plasmid was cultured in 50 ml LB medium containing 5 mg/l kanamycin in a 250-ml flask with one baffle, for 16 hours at 37°C, with shaking at 130 rpm. 0.5 ml of this culture was used to inoculate 500 ml LB medium with 5 mg/l kanamycin in a 2-liter flask with one baffle, and the

culture was incubated for 4 hours at 37°C followed by 4 hours and 30°C, with shaking at 130 rpm throughout. The cells were harvested by centrifugation, washed in 50 mM potassium phosphate buffer, and centrifuged again.

For the resting cell assays, 90 mg wet cells were weighed into deep-well plates in triplicate and resuspended in 0.5 ml 50 mM potassium phosphate buffer. For cell-free extracts, 4 grams wet cells in 8 ml disruption buffer were disrupted in French press.

For the resting cell assays, 5 µl of substrate (2.5 mg/ml in 2-propanol) was added to the cell suspension. The plate was sealed with air permeable foil, and the reaction was incubated on an orbital shaker at 1000 rpm at 28°C for 22 hours. No conversion of avermectin to 4"-ketoavermectin was detected.

For the cell-free assays, 100 µl cell free extract, 1µl substrate solution (20 mg/ml) in 2-propanol, 5 µl 100 mM NADPH, 10 µl ferredoxin, 10 µl ferredoxin reductase, and 374 µl potassium phosphate buffer pH 7.0 were added as described in Example III, and the assay was incubated at 30°C with shaking at 600 rpm for 20 hours. 9.2% +/- 0.3% of avermectin was converted to 4"-ketoavermectin.

Thus, expression of the *emaI* gene in *E. coli* resulted in the production of the active EmaI P450 monooxygenase enzyme which, when purified from the cells, was able to convert avermectin to 4"-ketoavermectin.

EXAMPLE XVI

Identification and Cloning of Genes Encoding Ferredoxin Reductases that Support Increased Activity of the $P450_{EmaI}$ Monooxygenase

The electron transport pathway that supports the activity of P450 monooxygenases also includes ferredoxin reductases. These proteins donate electrons to the ferredoxin and, as is the case with ferredoxins and P450 monooxygenases, specific ferredoxin reductases are known to be better electron donors for certain ferredoxins than others.

According, a number of ferredoxin reductase genes from *Streptomyces* strains were cloned and were evaluated for their impacts on the biocatalysis reaction. To do this, numerous bacterial ferredoxin reductase (Fre) protein sequences were retrieved from NCBI and aligned with the program Pretty from the GCG package. Two conserved regions,

approximately 266 amino acid residues apart, were used to make degenerate oligonucleotides for PCR. The forward primer (CGSCCSCCSCTSWSSAAS (SEQ ID NO:96; where "S" is C or G; and "W" is A or G)) and the reverse primer (SASSGCSTTSBCCCARTGYTC (SEQ ID NO:97; where "S" is C or G; "B" is C, G, or T; "R" is A or G; and "Y" is C or T)) were used to amplify 800 bp products from the biocatalytically active *Streptomyces* strains R-922 and I-1529. These pools of products were cloned into TOPO TA cloning vectors (commercially available from Invitrogen Inc., Carlsbad, CA), and 20 clones each from R922 and I-1529 were sequenced according to standard methods (see, *e.g.*, Current Protocols in Molecular Biology, eds. Ausubel *et al.*, John Wiley & Sons, Inc. 2000). Sequencing revealed that 4 unique *fre* gene fragments were isolated from the strains: three from R922 (*fre*3, *fre*12, *fre*14) and one from I-1529 (*fre*16). The *fre*3, *fre*12, *fre*14, and *fre*16 gene fragments were used as probes to identify full-length ferredoxin reductases from genomic clone banks of *Streptomyces* strains R922 and I-1529. By this approach, the complete coding sequence of each of the 4 different *fre* genes was cloned and sequenced. The nucleic acid and amino acid sequences are provided as follows: *fre*3 (SEQ ID NOs:98 and 99); *fre*12 (SEQ ID NOs:100 and 101); *fre*14 (SEQ ID NOs:102 and 103); and *fre*16 (SEQ ID NOs:104 and 105).

In order to assess the biological activity of each *fre* gene in relation to the activity of Ema1, each gene was inserted into the *ema1/fd233* operon described above, 3' to the *fd233* gene. This resulted in the formation of artificial operons consisting of the *ema1*, *fd233*, and individual *fre* genes that were expressed from the same promoter. The *ema1/fd233/fre* operons were cloned into the *Pseudomonas* plasmid pRK290 and introduced into 3 different *P. putida* strains. These strains were then analysed for Ema1 biocatalysis activity using the whole cell assay and one of the genes, the *fre* gene *fre*16 from strain I-1529, was found to increase the activity of $P450_{Ema1}$ monooxygenase by approximately 2-fold. This effect was strain specific, as it was seen only in one of the *P. putida* strains, ATCC Desposit No. 17453, and not in the other two. In *P. putida* strain ATCC 17453, the presence of *fre* gene *fre*16 resulted in 44% conversion of avermectin to 4"-keto-avermectin, as compared to 23% without this gene. The other *fre* genes had no impact on the biocatalysis activity in any of the *P. putida* strains tested.

In a similar approach, each of the *ema1/fd233/fre* operons were cloned into the *Streptomyces* plasmids pTUA, pTBBKA, and pEAA, and introduced into *S. lividans* strain

ZX7. In each case there was no impact in *S. lividans* by any of the *fre* genes on biocatalysis activity.

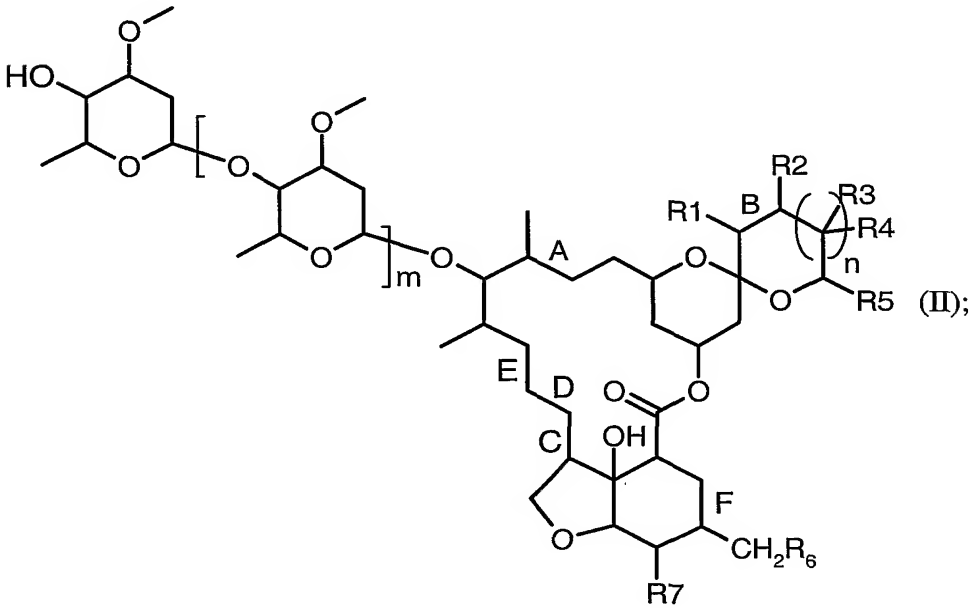
EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain, using no more than routine experimentation, numerous equivalents to the specific substances and procedures described herein. Such equivalents are considered to be within the scope of this invention.

The patent and scientific literature referred to herein establishes knowledge that is available to those with skill in the art. The issued patents, applications, and references, including GenBank database sequences, that are cited herein are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference. Any conflict between any reference cited herein and the specific teachings of this specification shall be resolved in favor of the latter. Likewise, any conflict between an art-understood definition of a word or phrase and a definition of the word or phrase as specifically taught in this specification shall be resolved in favor of the latter.

What is claimed is:

1. A purified nucleic acid molecule encoding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and is capable of regioselectively oxidizing the alcohol at position 4" of a compound of formular (II)



(II);

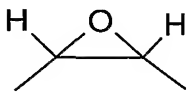
wherein

R_1-R_7 represent, independently of each other hydrogen or a substituent;

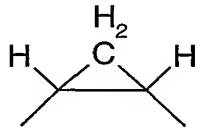
m is 0, 1 or 2;

n is 0, 1, 2 or 3; and

the bonds marked with A, B, C, D, E and F indicate, independently of each other, that two adjacent carbon atoms are connected by a double bond, a single bond, a single bond and a epoxide bridge of the formula



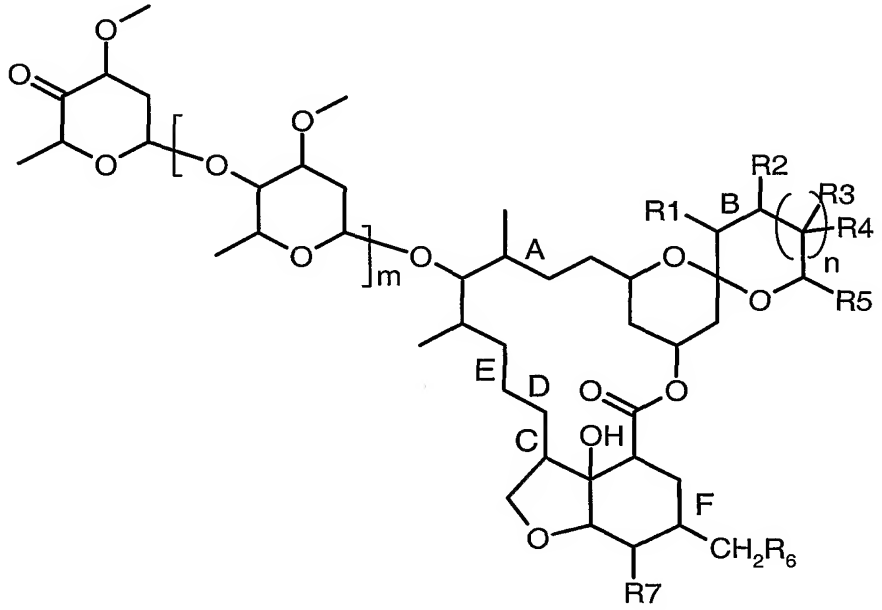
, or a single bond and a methylene bridge of the formula



,

including, where applicable, an E/Z isomer, a mixture of E/Z isomers, and/or a tautomer thereof, in each case in free form or in salt form,

in order to produce a compound of the formula (III)



(III),

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (II) above.

2. The nucleic acid molecule of claim 1, comprising a nucleic acid sequence that encodes a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"keto-avermectin.

3. The nucleic acid molecule of claims 1 or 2, comprising a nucleic acid sequence that encodes a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase, which polypeptide is substantially similar, and has between at least 50%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO:2.

4. The nucleic acid molecule of claim 3 comprising a nucleotide sequence
   a) as given in SEQ ID NO:1;
   b) having substantial similarity to (a);
   c) capable of hybridizing to (a) or the complement thereof;
   d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO:1, or the complement thereof;
   e) complementary to (a), (b) or (c);
   f) which is the reverse complement of (a), (b) or (c); or
   g) which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin. .

5. The nucleic acid molecule of claims 1 or 2, comprising a nucleic acid sequence that is at least 66 % identical to SEQ ID NO:1.

6. The nucleic acid molecule of claims 1 or 2, comprising a nucleic acid sequence that encodes a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase, which polypeptide is substantially similar, and has at least between 60%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.

7. The nucleic acid molecule of claims 1 or 2, comprising a nucleic acid sequence that encodes a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase,which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18,

SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, or SEQ ID NO:95.

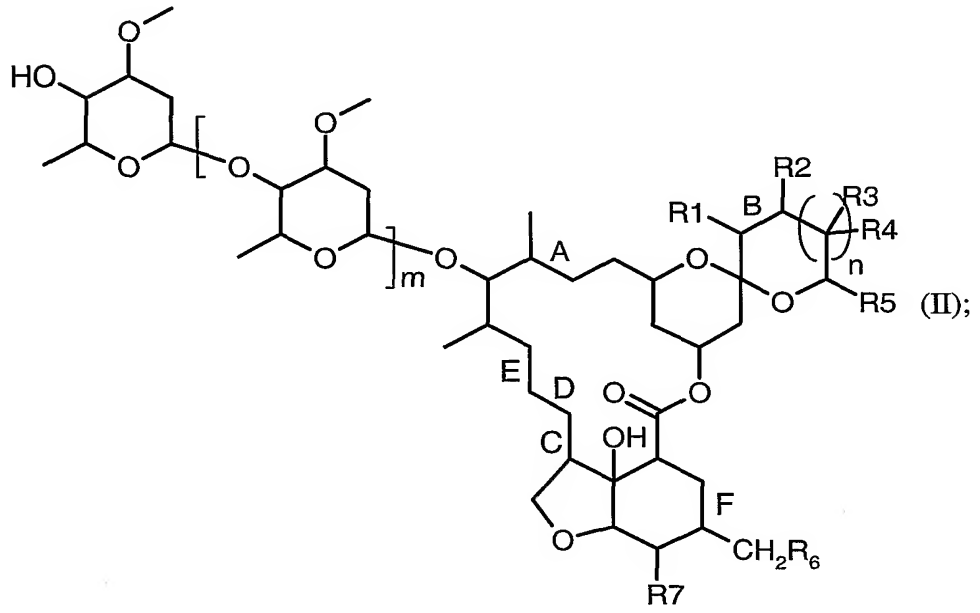
8. The nucleic acid molecule of claims 1 or 2 comprising a nucleotide sequence
   a) as given in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94;
   b) having substantial similarity to (a);
   c) capable of hybridizing to (a) or the complement thereof;
   d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94 or the complement thereof;
   e) complementary to (a), (b) or (c); and
   f) which is the reverse complement of (a), (b) or (c).
   g) . which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.

9. The nucleic acid molecule of claim 8, comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, and SEQ ID NO:94.

10. The nucleic acid molecule of anyone of claims 1 to 9, wherein the nucleic acid molecule is isolated from a *Streptomyces* strain.

11. The nucleic acid molecule of anyone of claims 1 to 10 further comprising a nucleic acid sequence encoding a tag which is linked to the P450 monooxygenase via a covalent bond.

12. A polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and is capable of regioselectively oxidizing the alcohol at position 4" of a compound of formular (II)



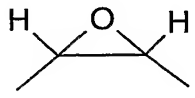
wherein

R_1-R_7 represent, independently of each other hydrogen or a substituent;

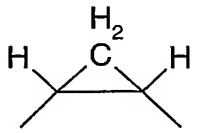
m is 0, 1 or 2;

n is 0, 1, 2 or 3; and

the bonds marked with A, B, C, D, E and F indicate, independently of each other, that two adjacent carbon atoms are connected by a double bond, a single bond, a single bond and a epoxide bridge of the formula



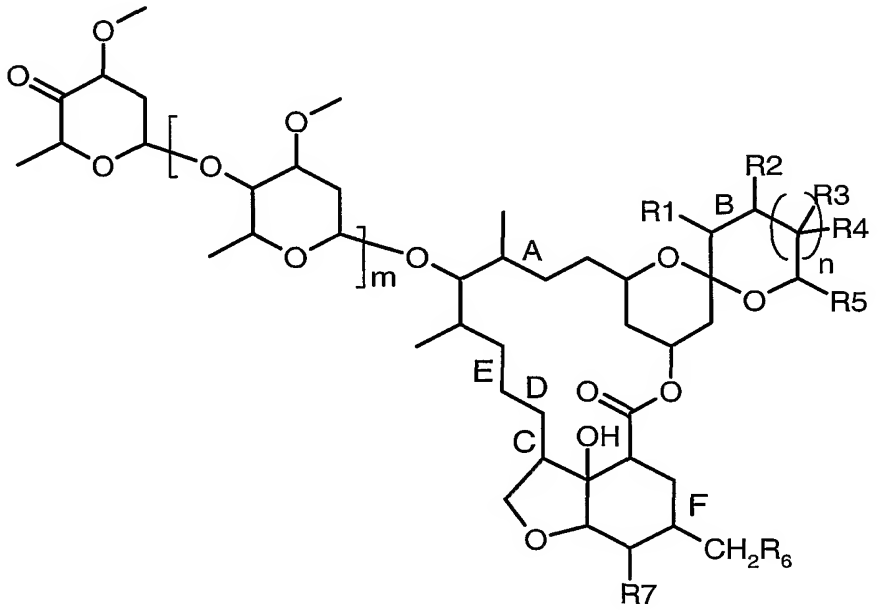
, or a single bond and a methylene bridge of the formula



,

including, where applicable, an E/Z isomer, a mixture of E/Z isomers, and/or a tautomer thereof, in each case in free form or in salt form,

in order to produce a compound of the formula (III)



(III),

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (II) above.

13. A polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"keto-avermectin.

14. The polypeptide of claims 12 or 13 that comprises an amino acid sequence that is encoded by a nucleic acid molecule

a) as given in SEQ ID NO:1 or the complement thereof;

b) having substantial similarity to (a);

c) capable of hybridizing to (a) or the complement thereof;

d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO:1, or the complement thereof;

e) complementary to (a), (b) or (c);

f) which is the reverse complement of (a), (b) or (c); or

g) which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.

15. The polypeptide of claims 12 to 14, comprising an amino acid sequence that is at least 50% identical to SEQ ID NO:2.

16. The polypeptide of claims 12 or 13 comprising an amino acid sequence that is encoded by a nucleic acid molecule

a) as given in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94 or the complement thereof;

b) having substantial similarity to (a);

c) capable of hybridizing to (a) or the complement thereof;

d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, or SEQ ID NO:94 or the complement thereof, or the complement thereof;

e) complementary to (a), (b) or (c);

f) which is the reverse complement of (a), (b) or (c); or

g) which is a functional part of (a), (b), (c), (d), (e) or (f) encoding a polypeptide that still exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"-keto-avermectin.

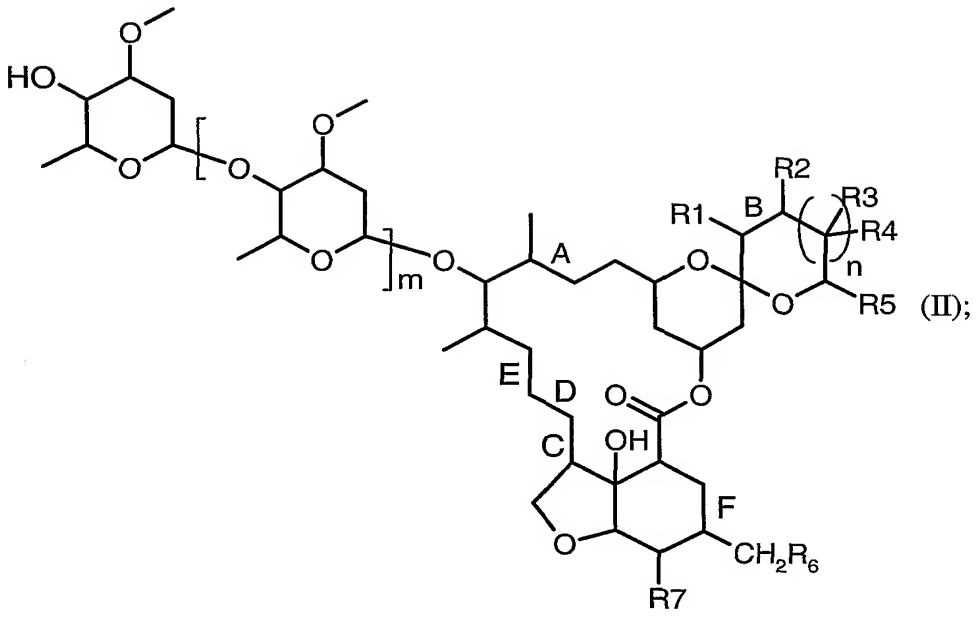
17. The polypeptide of claim 16, comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, and SEQ ID NO:95.

18. The polypeptide of anyone of claims 12 to 17, further comprising a tag.

19. A binding agent that specifically binds to the polypeptide of anyone of claims 12 to 18.

20. The binding agent of claim 20, wherein the binding agent is an antibody.

21. A family of polypeptides exhibiting an enzymatic activity of a P450 monooxygenase, wherein each member of the family is capable of regioselectively oxidizing the alcohol at position 4" of a compound of formular (II)



(II);

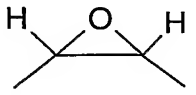
wherein

R_1-R_7 represent, independently of each other hydrogen or a substituent;

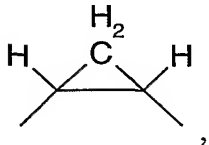
m is 0, 1 or 2;

n is 0, 1, 2 or 3; and

the bonds marked with A, B, C, D, E and F indicate, independently of each other, that two adjacent carbon atoms are connected by a double bond, a single bond, a single bond and a epoxide bridge of the formula



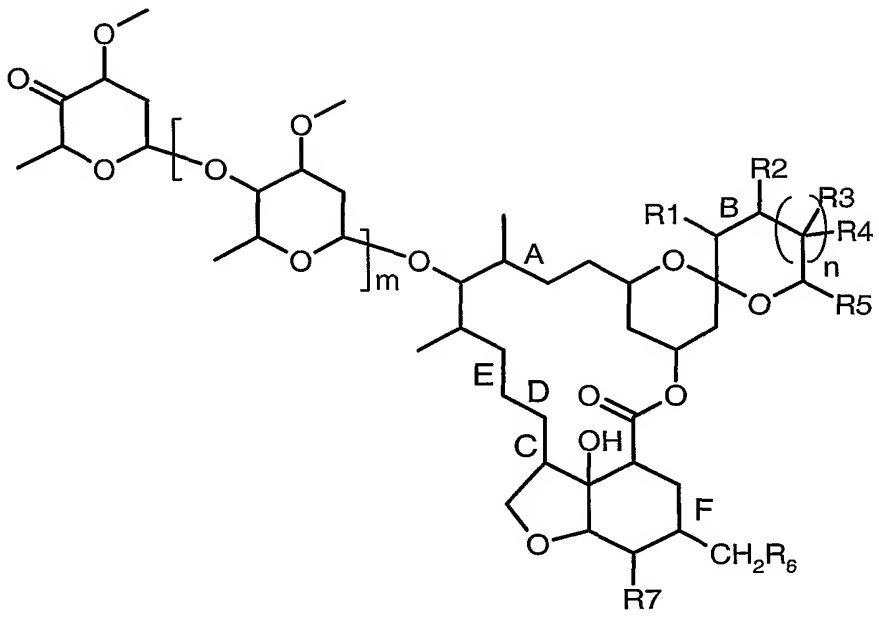
, or a single bond and a methylene bridge of the formula



,

including, where applicable, an E/Z isomer, a mixture of E/Z isomers, and/or a tautomer thereof, in each case in free form or in salt form,

in order to produce a compound of the formula (III)



(III),

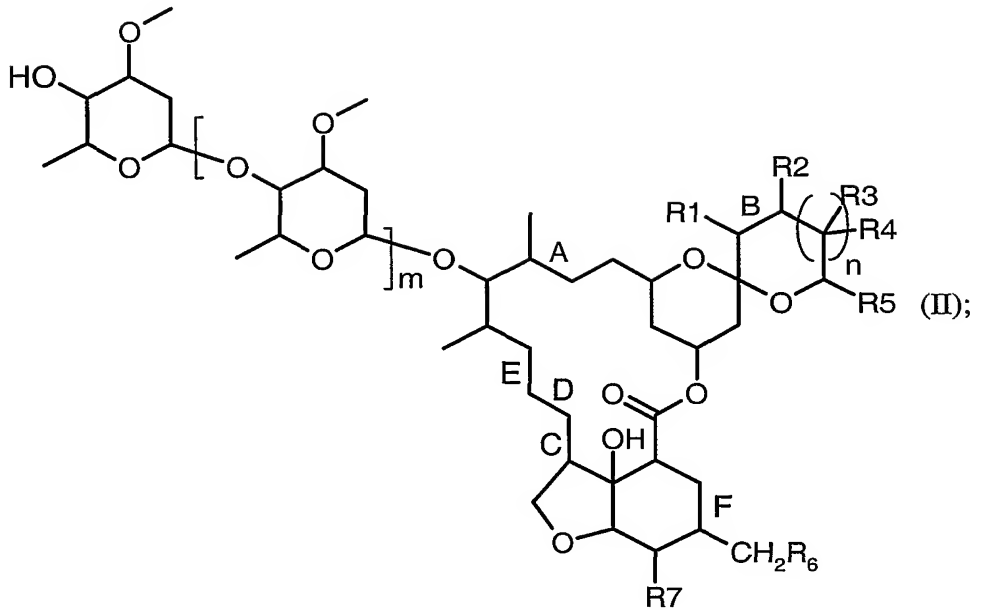
wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (II) above.

22. A family of polypeptides exhibiting an enzymatic activity of a P450 monooxygenase, wherein each member of the family oxidizes avermectin to 4"keto-avermectin.

23. The family of claims 21 or 22, wherein each member of the family is comprises an amino acid sequence that is at least 50% identical to SEQ ID NO:2.

24. A purified nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide exhibiting an enzymatic activity of a ferredoxin and a ferredoxin reductase,respectively, wherein the nucleic acid molecule is isolated from a *Streptomyces* strain comprising a P450 monooxygenase that is capable of regioselectively oxidizing the alcohol at position 4" of a compound of formular (II)



(II);

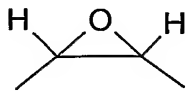
wherein

R_1-R_7 represent, independently of each other hydrogen or a substituent;

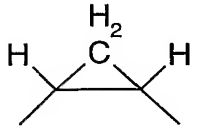
m is 0, 1 or 2;

n is 0, 1, 2 or 3; and

the bonds marked with A, B, C, D, E and F indicate, independently of each other, that two adjacent carbon atoms are connected by a double bond, a single bond, a single bond and a epoxide bridge of the formula



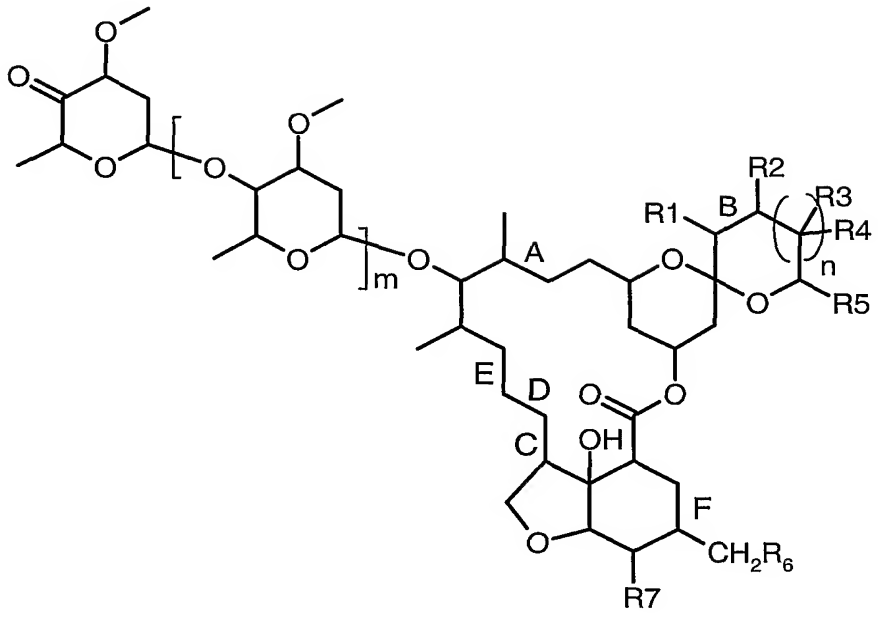
, or a single bond and a methylene bridge of the formula



,

including, where applicable, an E/Z isomer, a mixture of E/Z isomers, and/or a tautomer thereof, in each case in free form or in salt form,

in order to produce a compound of the formula (III)



(III),

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (II) above.

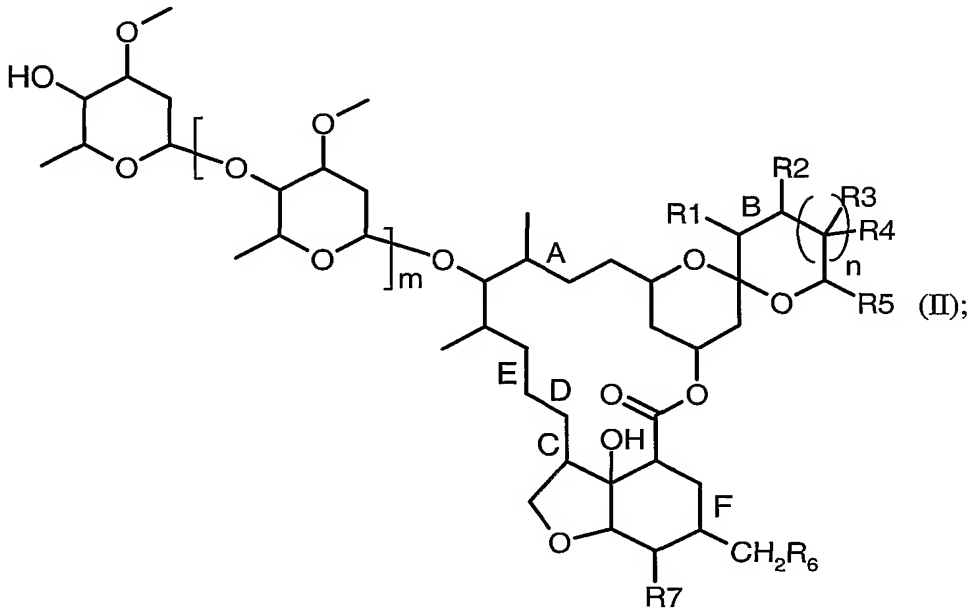
25. A purified nucleic acid molecule according to claim 24 comprising a nucleotide sequence encoding a polypeptide exhibiting an enzymatic activity of a ferredoxin and a ferredoxin reductase, respectively, wherein the nucleic acid molecule is isolated from a *Streptomyces* strain comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"keto-avermectin.

26. The nucleic acid molecule of claim 25, comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:35 and SEQ ID NO:37.

27. The nucleic acid molecule of claim 25, comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:102, and SEQ ID NO:104.

28. A polypeptide exhibiting an enzymatic activity of a ferredoxin and a ferredoxin reductase, respectively, wherein the polypeptide is isolated from a *Streptomyces* strain comprising a

P450 monooxygenase that is capable of regioselectively oxidizing the alcohol at position 4" of a compound of formular (II)



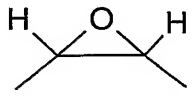
wherein

R_1-R_7 represent, independently of each other hydrogen or a substituent;

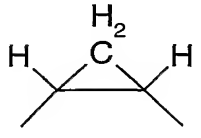
m is 0, 1 or 2;

n is 0, 1, 2 or 3; and

the bonds marked with A, B, C, D, E and F indicate, independently of each other, that two adjacent carbon atoms are connected by a double bond, a single bond, a single bond and a epoxide bridge of the formula

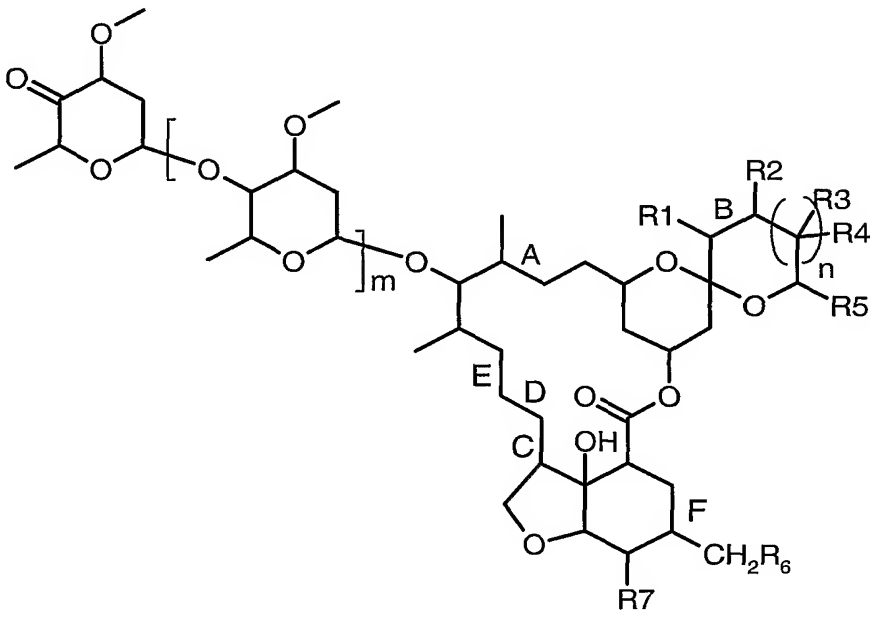


, or a single bond and a methylene bridge of the formula



,

including, where applicable, an E/Z isomer, a mixture of E/Z isomers, and/or a tautomer thereof, in each case in free form or in salt form,
in order to produce a compound of the formula (III)



(III),

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (II) above.

29. A polypeptide exhibiting an enzymatic activity of a ferredoxin and a ferredoxin reductase, respectively, wherein the ferredoxin protein is isolated from a *Streptomyces* strain comprising a P450 monooxygenase that regioselectively oxidizes avermectin to 4"keto-avermectin.

30. The ferredoxin protein of claim 29, comprising an amino acid sequence selected from the group consisting of SEQ ID NO:36 and SEQ ID NO:38.

31. The ferredoxin reductase protein of claim 29, comprising an amino acid sequence selected from the group consisting of SEQ ID NO:99, SEQ ID NO:101, SEQ ID NO:103, and SEQ ID NO:105.

32. A cell genetically engineered to comprise a nucleic acid molecule encoding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase according to anyone of claims 1 to 11.

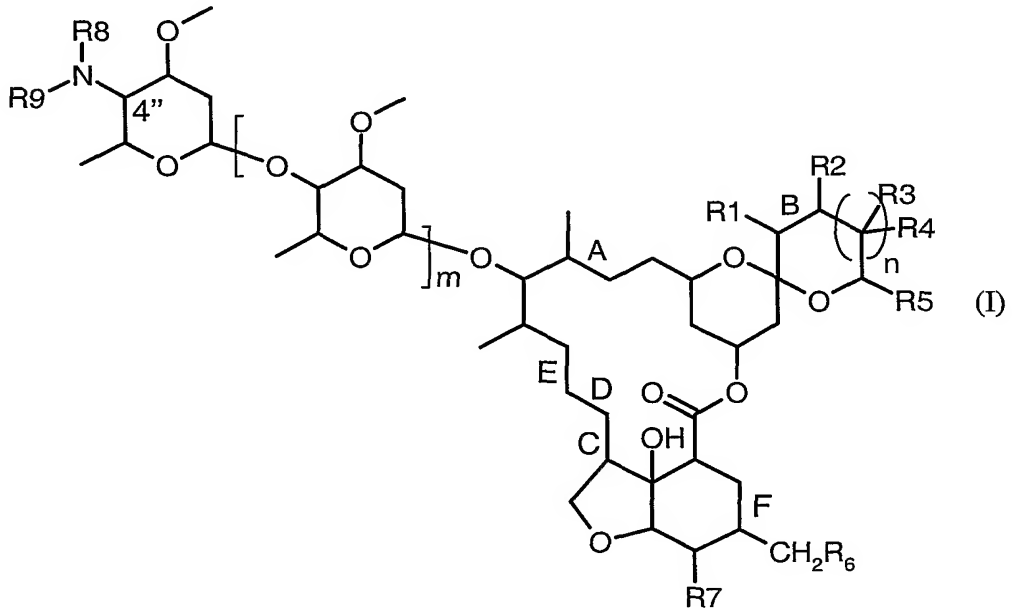
33. The cell of claim 32 further comprising a nucleic acid molecule encoding a ferredoxin protein and a ferredoxin reductrase protein, respectively, or a combination thereof.

34. The cell of claims 32 or 33, wherein the nucleic acid molecule is positioned for expression in the cell.

35. The cell of anyone of claims 32 to 34, wherein the cell is a genetically engineered cell selected from the group consisting of a *Streptomyces* strain cell and a *Pseudomona s*train cell, and an *Escherichia coli* strain cell.

36. The cell of claim 35, wherein the cell has NRRL Designation No. B-30478 and NRRL Designation No.B-30479, respectively.

37. A method for the preparation a compound of the formula



(I)

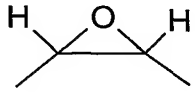
in which

R_1-R_9 represent, independently of each other hydrogen or a substituent;

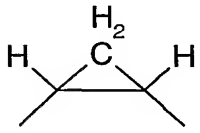
m is 0, 1 or 2;

n is 0, 1, 2 or 3; and

the bonds marked with A, B, C, D, E and F indicate, independently of each other, that two adjacent carbon atoms are connected by a double bond, a single bond, a single bond and a epoxide bridge of the formula



, or a single bond and a methylene bridge of the formula

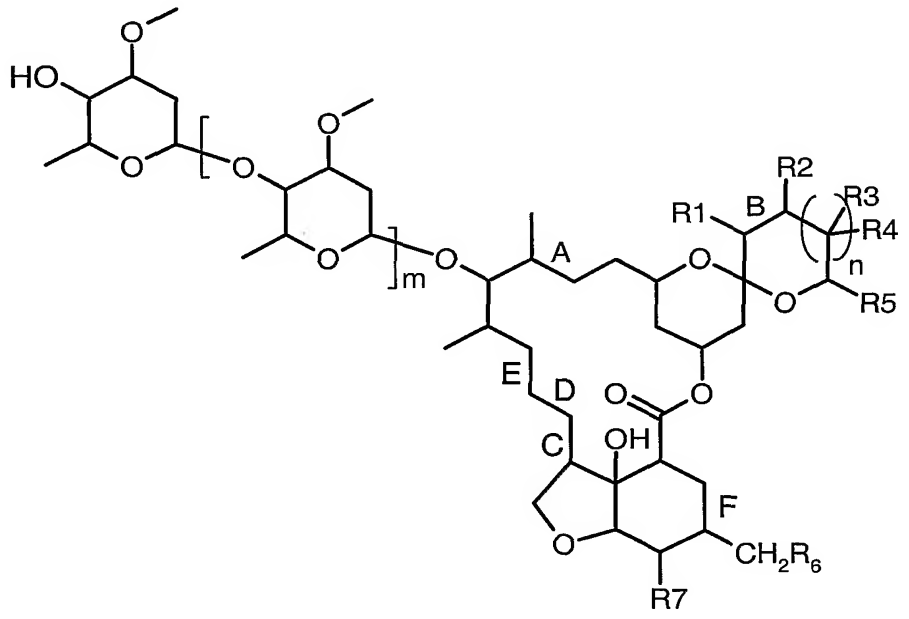


,

including, where applicable, an E/Z isomer, a mixture of E/Z isomers, and/or a tautomer thereof, in each case in free form or in salt form,

which process comprises

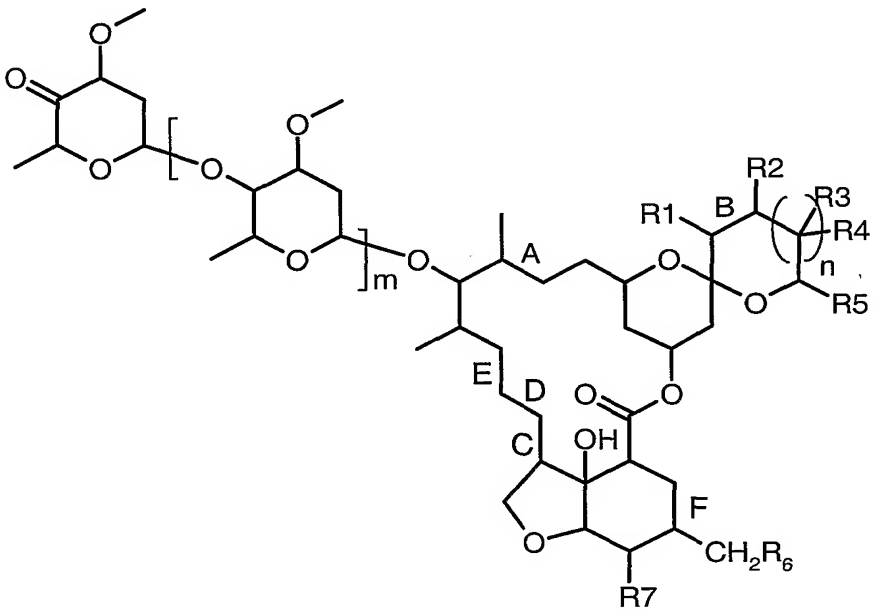
1) bringing a compound of the formula



(II);

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (I) above, into contact with a polypeptide according to the invention that is capable of regioselectively oxidising the alcohol at position 4" in order to form a compound of the formula



(III),

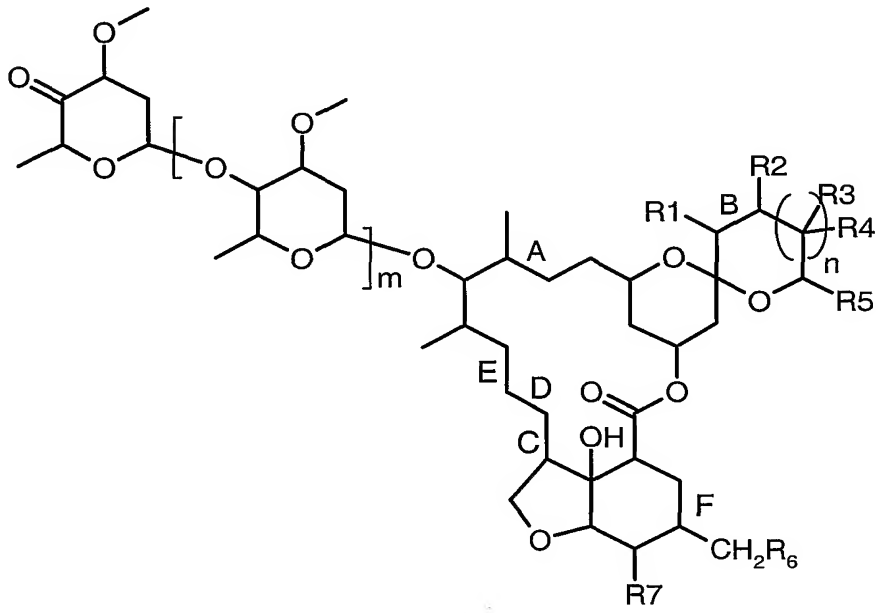
in which R_1, R_2, R_3, R_4, R_5, R_6, R_7, m, n, A, B, C, D, E and F have the meanings given for formula (I); and

2) reacting the compound of the formula (III) with an amine of the formula $HN(R_8)R_9$, wherein R_8 and R_9 have the same meanings as given for formula (I), and which is known, in the presence of a reducing agent;

and, in each case, if desired, converting a compound of formula (I) obtainable in accordance with the process or by another method, or an E/Z isomer or tautomer thereof, in each case in free form or in salt form, into a different compound of formula (I) or an E/Z isomer or tautomer thereof, in each case in free form or in salt form, separating a mixture of E/Z isomers obtainable in accordance with the process and isolating the desired isomer, and/or converting a free compound of formula (I) obtainable in accordance with the process or by another method, or an E/Z isomer or tautomer thereof, into a salt or converting a salt, obtainable in accordance with the process or by another

method, of a compound of formula (I) or of an E/Z isomer or tautomer thereof into the free compound of formula (I) or an E/Z isomer or tautomer thereof or into a different salt.

38. A method for the preparation of a compound of the formula

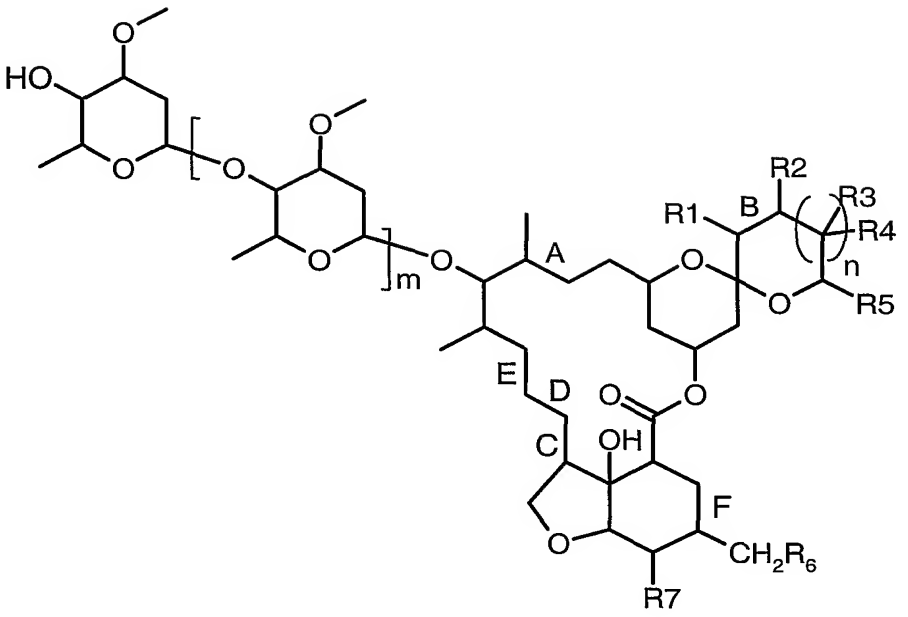


(III),

in which R_1, R_2, R_3, R_4, R_5, R_6, R_7, m, n, A, B, C, D, E and F have the meanings given for formula (III) of claim 37,

which process comprises

1) bringing a compound of the formula



(II);

wherein

R_1-R_7, m, n, A, B, C, D, E and F have the same meanings as given for formula (I) above, into contact with a polypeptide according to the invention that is capable of regioselectively oxidising the alcohol at position 4", maintaining said contact for a time sufficient for the oxidation reaction to occur and isolating and purifying the compound of formula (II).

39. A method according to anyone of claims 37 or 38 for making emamectin, comprising adding a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"keto-avermectin to a reaction mixture comprising avermectin and incubating the reaction mixture under conditions that allow the polypeptide to regioselectively oxidize avermectin to 4"keto-avermectin.

40. The method of anyone of claims 37 to 39, wherein the reaction mixture further comprises a ferredoxin protein.

41. The method of anyone of claims 37 to 40, wherein the reaction mixture further comprises a ferredoxin reductase protein.

42. A formulation for making emamectin comprising a polypeptide that exhibits an enzymatic activity of a P450 monooxygenase and regioselectively oxidizes avermectin to 4"keto-avermectin.

5

43. The formulation of claims 42 further comprising a ferredoxin protein.

44. The formulation of claim 42 or 43 further comprising a ferredoxin reductase protein.

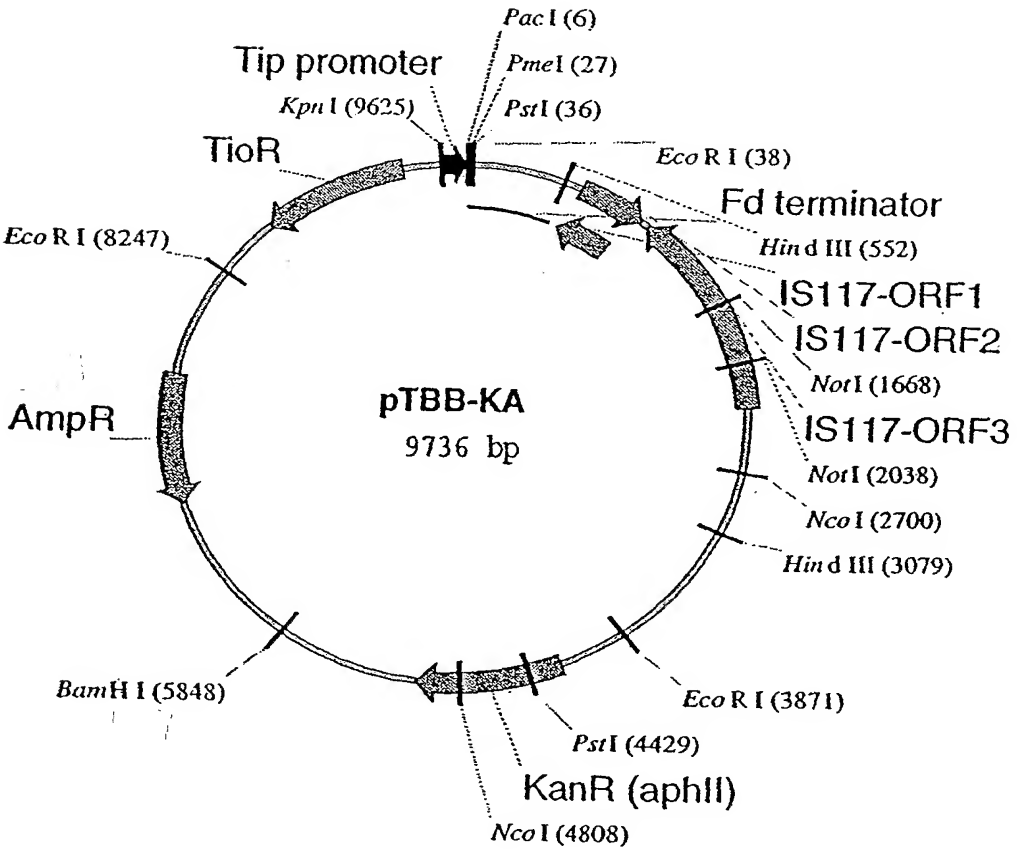

Pac I (6)
Tip promoter
Pme I (27)
Kpn I (9625)
Pst I (36)
TioR
Eco R I (38)
Fd terminator
Eco R I (8247)
Hin d III (552)
IS117-ORF1
IS117-ORF2
pTBB-KA
9736 bp
Not I (1668)
AmpR
IS117-ORF3
Not I (2038)
Nco I (2700)
Hin d III (3079)
BamH I (5848)
Eco R I (3871)
Pst I (4429)
KanR (aphII)
Nco I (4808)

Figure 1

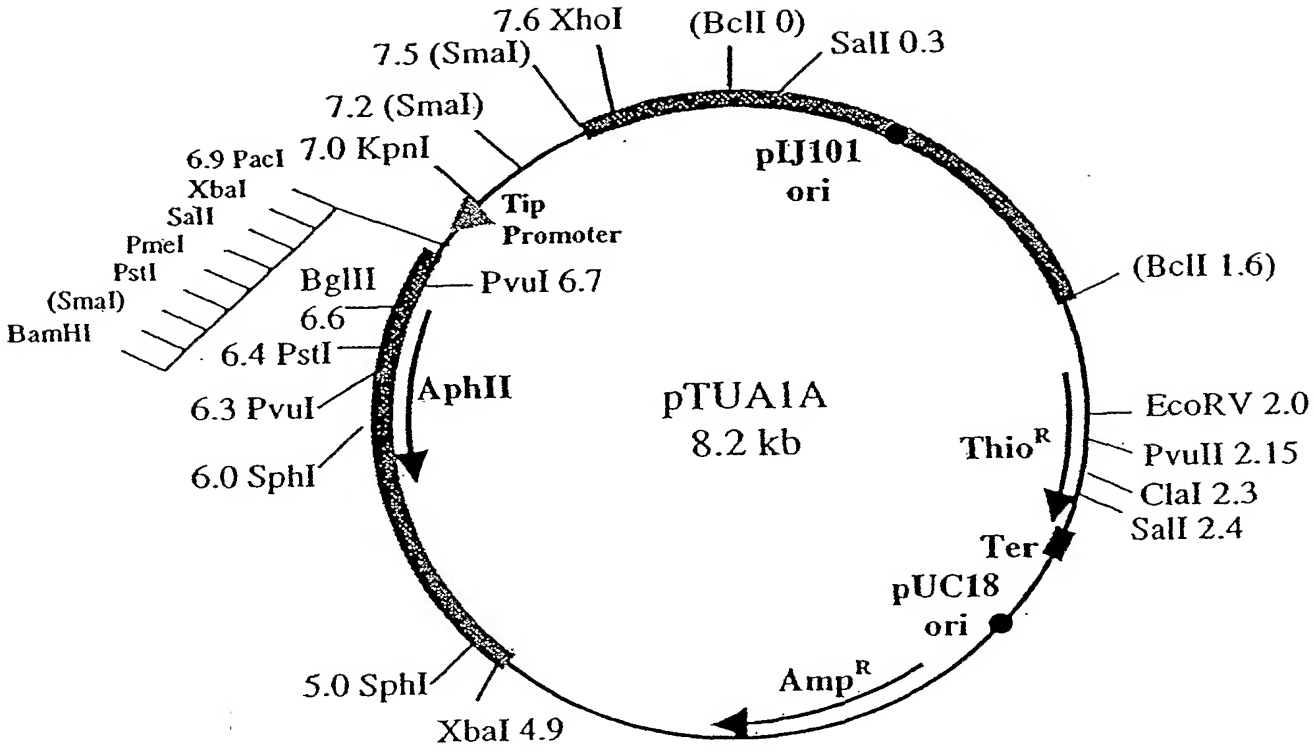

7.6 XhoI
(BclI 0)
SalI 0.3
7.5 (SmaI)
7.2 (SmaI)
7.0 KpnI
6.9 PacI
XbaI
SalI
PmeI
PstI
(SmaI)
BamHI
Tip
Promoter
pIJ101
ori
(BclI 1.6)
BglII
PvuI 6.7
6.6
6.4 PstI
6.3 PvuI
6.0 SphI
AphII
pTUA1A
8.2 kb
EcoRV 2.0
Thio^R
PvuII 2.15
ClaI 2.3
SalI 2.4
Ter
pUC18
ori
Amp^R
5.0 SphI
XbaI 4.9

Figure 2

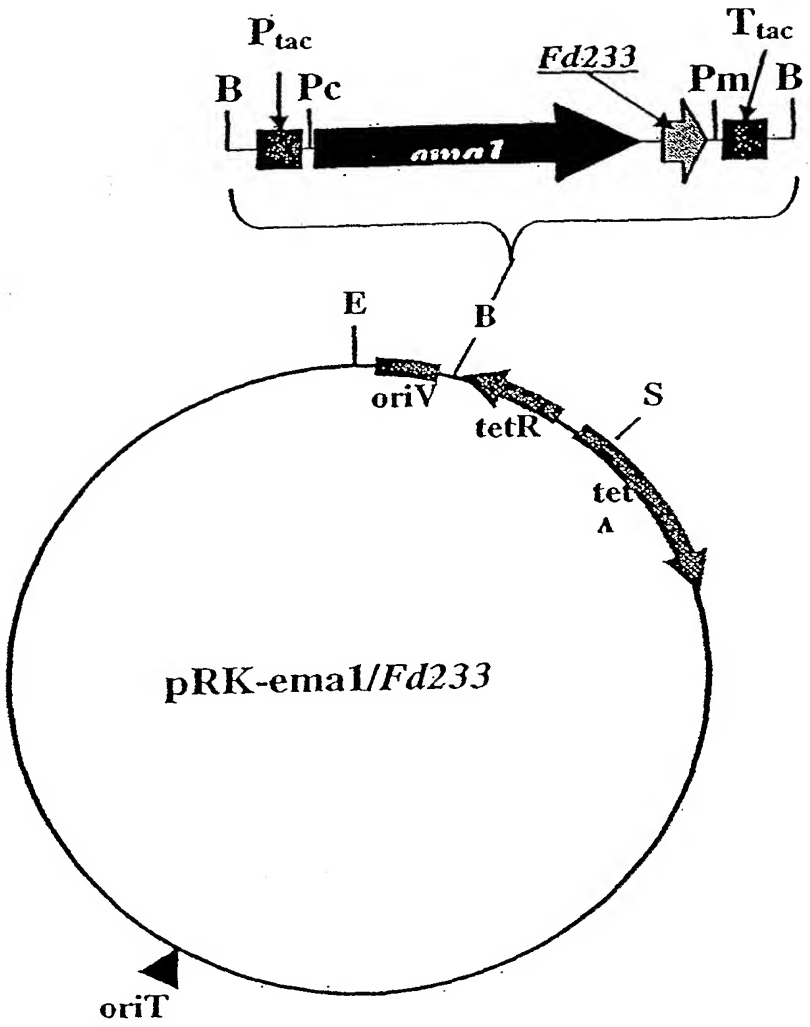

Ptac
B
Pc
ema1
Fd233
Pm
B
Ttac
E
B
oriV
tetR
S
tet
A
pRK-ema1/Fd233
oriT

Figure 3

SEQUENCE LISTING

<110> Syngenta Participations AG

<120> METHODS AND COMPOSITIONS FOR MAKING EMAMECTIN

<130> PB/5-60016A

<140>
<141>

<150> US 60/291,149
<151> 2001-05-16

<160> 105

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1293
<212> DNA
<213> Streptomyces tubercidicus

<400> 1
atgtcggaat taatgaactc tccgttcgcc gcgcacgtcg ggaaacaccc gggcgagccg 60
aatgtgatgg accccgccct gatcaccgac ccgttcaccg gctacggcgc gctgcgtgag 120
cagggcccgg tcgtacgggg ccggttcatg gacgactcgc ccgtctggct ggtgacgcgg 180
ttcgaggagg tccgccaggt cctgcgcgac cagcggttcg tgaacaatcc ggcctcgccg 240
tccctgaact acgcgcccga ggacaacccg ctgacccggc tgatggagat gctgggcctc 300
cccgagcacc tccgcgtcta cctgctcgga tcgatcctca actacgacgc ccccgaccac 360
acccggctgc gccgtctggt gtcgcgggcg ttcacggccc gcaagatcac cgacctgcgg 420
ccccgggtcg agcagatcgc cgacgcgctg ctggcccggc tgcccgagca cgccgaggac 480
ggcgtcgtcg acctcatcca gcacttcgcc tacccccctgc cgatcaccgt catctgcgaa 540
ctggtcggca tacccgaagc ggaccgcccg cagtggcgaa cgtggggcgc cgacctcatc 600
tcgatggatc cggaccggct cggcgcctcg ttcccggcga tgatcgagca catccatcag 660
atggtccggg aacggcgcga ggcgctcacc gacgacctgc tcagcgaact gatccgcacc 720
catgacgacg acggcgggcg gctcagcgac gtcgagatgg tcaccatgat cctcacgctc 780
gtcctcgccg gccacgagac caccgcccac ctcatcagca acggcacggc ggcgctgctc 840
acccaccccg accagctgcg tctggtcaag gacgatccgg ccctcctccc ccgtgccgtc 900
cacgagctga tgcgctggtg cgggccggtg cacatgaccc agctgcgcta cgccaccgcc 960
gacgtcgacc tcgccggcac accgatccgc cagggcgatg ccgttcaact catcctggta 1020
tcggccaact tcgacccccg tcactacacc gaccccgacc gcctcgatct cacccggcac 1080
cccgcgggcc acgccgagaa ccatgtgggt ttcggccatg gagcgcacta ctgcctgggc 1140
gccacactcg ccaaacagga aggtgaagtc gccttcggca aactgctcac gcactacccg 1200
gacatatcgc tgggcatcgc cccggaacac ctggagcgga caccgctgcc gggcaactgg 1260
cggctgaact cgctgccggt gcggttgggg tga 1293

<210> 2
<211> 430
<212> PRT
<213> Streptomyces tubercidicus

<400> 2

```
Met Ser Glu Leu Met Asn Ser Pro Phe Ala Ala His Val Gly Lys His
1               5                  10                  15
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Thr Asp Pro Phe
            20                  25                  30
Thr Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Met Asp Asp Ser Pro Val Trp Leu Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Gln Val Leu Arg Asp Gln Arg Phe Val Asn Asn Pro Ala Ser Pro
65                  70                  75                  80
Ser Leu Asn Tyr Ala Pro Glu Asp Asn Pro Leu Thr Arg Leu Met Glu
                85                  90                  95
Met Leu Gly Leu Pro Glu His Leu Arg Val Tyr Leu Leu Gly Ser Ile
            100                 105                 110
Leu Asn Tyr Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu
    130                 135                 140
Gln Ile Ala Asp Ala Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Ala Asp Arg Pro Gln Trp
            180                 185                 190
Arg Thr Trp Gly Ala Asp Leu Ile Ser Met Asp Pro Asp Arg Leu Gly
        195                 200                 205
Ala Ser Phe Pro Ala Met Ile Glu His Ile His Gln Met Val Arg Glu
    210                 215                 220
Arg Arg Glu Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Gly Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Ile Leu Thr Leu Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Ser Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Val Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val His Met Thr Gln Leu Arg Tyr Ala Thr Ala
305                 310                 315                 320
Asp Val Asp Leu Ala Gly Thr Pro Ile Arg Gln Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Phe Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Phe Gly Lys Leu Leu Thr His Tyr Pro
385                 390                 395                 400
Asp Ile Ser Leu Gly Ile Ala Pro Glu His Leu Glu Arg Thr Pro Leu
                405                 410                 415
Pro Gly Asn Trp Arg Leu Asn Ser Leu Pro Val Arg Leu Gly
```

420 425 430

<210> 3
<211> 1293
<212> DNA
<213> Streptomyces tubercidicus

<400> 3
atgtcggcat tatccagctc tccgttcgct gcgcatgtcg ggaaacaccc gggtgagccg 60
aatgtgatgg agccggcgct gctcaccgac ccgttcgcgg gctacggcgc gctgcgtgag 120
caggccccgg tcgtacgggg ccggttcgtg gacgactcac cggtctggtt cgtgacgcgc 180
ttcgaggagg tccgccaagt cctgcgcgac cagcggttcg tgaacaatcc ggccgcgccg 240
cccctggccc catcggccga ggagaacccg ctgaccaggc tgatggacat gctgggcctc 300
cccgagcacc tccgcgtcta catgctcggg tcgattctca actacgacgc ccccgaccac 360
acccggctgc gccgtctggt gtcgcgcgcg ttcacggcgc ggaagatcac cgatctgcga 420
ccgcgtgtcg agcagatcgc cgacgagctg ctggcccgcc tccccgagta cgccgaggac 480
ggcgtcgtcg acctcatcca gcatttcgcc tacccgctgc cgatcaccgt catctgcgag 540
ctggtcggca tacccgaagc ggaccgcccg cagtggcgga agtggggcgc cgacctcatc 600
tcgatggacc cggaccggct cggcgcaacg ttcccggcga tgatcgagca catccatgag 660
atggtccggg agcggcgcgc ggcgctcacc gatgatctgc tcagcgagct gatccgtacc 720
catgacgacg atggcggccg gctcagcgac gtcgagatgg tcaccatgat cctcacgctc 780
gtcctcgccg gtcacgagac caccgcccac ctcatcagca acggcacggc ggcgctgctc 840
acccaccccg accagctgcg cctgctcaag gacgacccgg ccctgctccc ccgggccgtc 900
catgaactga tgcgctggtg cgggccggtg cagatgacgc agctgcgcta cgcggccgcc 960
gacgtcgacc tcgccggtac gcggatccac aagggcgacg ccgtacaact cctcctggtt 1020
gcggcgaact tcgacccccg ccactacacc gaccccgacc gtctcgatct gacgcgtcac 1080
cccgccggcc acgccgagaa ccatgtgggt ttcggccacg gtgcgcatta ctgcctgggt 1140
gccaccctcg ccaagcagga gggcgaagtc gcgttcggca agctgctcgc gcactacccg 1200
gagatgtccc tgggcatcga accggaacgt ctggagcgat tgccgctgcc tggcaactgg 1260
cggctgaatt ccctgccgtt gcggctgggg tga 1293

<210> 4
<211> 430
<212> PRT
<213> Streptomyces tubercidicus

<400> 4
Met Ser Ala Leu Ser Ser Ser Pro Phe Ala Ala His Val Gly Lys His
1 5 10 15
Pro Gly Glu Pro Asn Val Met Glu Pro Ala Leu Leu Thr Asp Pro Phe
20 25 30
Ala Gly Tyr Gly Ala Leu Arg Glu Gln Ala Pro Val Val Arg Gly Arg
35 40 45
Phe Val Asp Asp Ser Pro Val Trp Phe Val Thr Arg Phe Glu Glu Val
50 55 60
Arg Gln Val Leu Arg Asp Gln Arg Phe Val Asn Asn Pro Ala Ala Pro
65 70 75 80
Pro Leu Ala Pro Ser Ala Glu Glu Asn Pro Leu Thr Arg Leu Met Asp
85 90 95
Met Leu Gly Leu Pro Glu His Leu Arg Val Tyr Met Leu Gly Ser Ile
100 105 110
Leu Asn Tyr Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser

```
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu
    130                 135                 140
Gln Ile Ala Asp Glu Leu Leu Ala Arg Leu Pro Glu Tyr Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Ala Asp Arg Pro Gln Trp
            180                 185                 190
Arg Lys Trp Gly Ala Asp Leu Ile Ser Met Asp Pro Asp Arg Leu Gly
        195                 200                 205
Ala Thr Phe Pro Ala Met Ile Glu His Ile His Glu Met Val Arg Glu
    210                 215                 220
Arg Arg Ala Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Gly Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Ile Leu Thr Leu Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Ser Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val Gln Met Thr Gln Leu Arg Tyr Ala Ala Ala
305                 310                 315                 320
Asp Val Asp Leu Ala Gly Thr Arg Ile His Lys Gly Asp Ala Val Gln
                325                 330                 335
Leu Leu Leu Val Ala Ala Asn Phe Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Phe Gly Lys Leu Leu Ala His Tyr Pro
385                 390                 395                 400
Glu Met Ser Leu Gly Ile Glu Pro Glu Arg Leu Glu Arg Leu Pro Leu
                405                 410                 415
Pro Gly Asn Trp Arg Leu Asn Ser Leu Pro Leu Arg Leu Gly
            420                 425                 430
```

<210> 5
<211> 1413
<212> DNA
<213> Streptomyces rimosus

<400> 5

```
atgaccacat cgcccaccga gtcccgggcg gccaccccgc ccgactccac cgcctccccc     60
tcgaccgctt ccgccccggc caccaccgct tcggccgccg cctctccgga caccaccgac    120
cgcaccacgc tcccctccta cgtcggcctc cacccgggcg agccgaacct gatggaaccg    180
gagctgctgg agaacccgta caccggctac ggcacgctgc gcgagcaggc cccgctcgtc    240
cgcgcccggt tcatcgacga ctcgcccatc tggctggtga cccgcttcga cgtggtgcgc    300
gaggtgatgc gtgaccagcg gttcgtcaac aacccgaccc tggtgcccgg catcggcgcg    360
gacaaggacc cgcgtgcccg gctgatcgag ctgttcggca tccccgagga cctggccccg    420
```

```
tacctcaccg acaacatcct caccagcgac ccgccggacc acacccggct gcgccgcctg    480
gtctcccgcg ccttcaccgc acgccgtatc caggacctgc ggccgcgcgt cgagcggatc    540
accgacgagc tgctggaacg gctgccggac catgccgagg acggcgtcgt cgacctcgtc    600
gagcacttcg cctacccgct gcccatcacg gtcatctgcg agctggtcgg catcgacgag    660
gaggatcggg cgctgtggcg gcggttcggc gccgacctcg cctcgctgaa ccccaagcgc    720
atcggcgcca ccatgccgga gatgatctcg cacatccacg agctgatcga cgaacggcgc    780
gcggccctgc gggacgacct gctcagcggg ctcatccggg cgcaggacga cgacggcggc    840
cggctgagcg acgtcgagat ggtcaccctg gtcctgaccc tggtactggc cggtcacgag    900
accaccgccc acctcatcag caacggcacc ctcgccctgc tcacccaccc cgaccagcgg    960
cggctgatcg acgaggaccc ggcgctgctg ccgcgcgcgg tccacgagct gatgcgctgg   1020
tgcgggccga tccaggccac ccagcttcgg tacgccctgg aggacaccga ggtggccgga   1080
gtccaggtcc gccagggcga ggccctgatg ttcagcctcg tcgcggccaa ccacgacccg   1140
cgccactaca ccgggccgga gcggctcgac ctgacgcggc agccggccgg ccgcgccgag   1200
gaccacgtcg gcttcggcca cggcatgcac tactgcctgg gtgcctcact cgcccggcag   1260
gaggccgagg tggcctacgg gaagctgctc acccgctacc cggacctggc gctcgccctc   1320
accccggaac agttggagga ccaggaacgc ctgcggcagc ccggcacctg gcgcctgcga   1380
cggctgccgc tgaggctgca cgcgcagagc tga                                1413
```

<210> 6
<211> 470
<212> PRT
<213> Streptomyces rimosus

<400> 6

```
Met Thr Thr Ser Pro Thr Glu Ser Arg Ala Ala Thr Pro Pro Asp Ser
1               5                  10                  15
Thr Ala Ser Pro Ser Thr Ala Ser Ala Pro Ala Thr Thr Pro Ser Ala
            20                  25                  30
Ala Ala Ser Pro Asp Thr Thr Asp Arg Thr Thr Leu Pro Ser Tyr Val
        35                  40                  45
Gly Leu His Pro Gly Glu Pro Asn Leu Met Glu Pro Glu Leu Leu Glu
    50                  55                  60
Asn Pro Tyr Thr Gly Tyr Gly Thr Leu Arg Glu Gln Ala Pro Leu Val
65                  70                  75                  80
Arg Ala Arg Phe Ile Asp Asp Ser Pro Ile Trp Leu Val Thr Arg Phe
                85                  90                  95
Asp Val Val Arg Glu Val Met Arg Asp Gln Arg Phe Val Asn Asn Pro
            100                 105                 110
Thr Leu Val Pro Gly Ile Gly Ala Asp Lys Asp Pro Arg Ala Arg Leu
        115                 120                 125
Ile Glu Leu Phe Gly Ile Pro Glu Asp Leu Ala Pro Tyr Leu Thr Asp
    130                 135                 140
Asn Ile Leu Thr Ser Asp Pro Pro Asp His Thr Arg Leu Arg Arg Leu
145                 150                 155                 160
Val Ser Arg Ala Phe Thr Ala Arg Arg Ile Gln Asp Leu Arg Pro Arg
                165                 170                 175
Val Glu Arg Ile Thr Asp Glu Leu Leu Glu Arg Leu Pro Asp His Ala
            180                 185                 190
Glu Asp Gly Val Val Asp Leu Val Glu His Phe Ala Tyr Pro Leu Pro
        195                 200                 205
Ile Thr Val Ile Cys Glu Leu Val Gly Ile Asp Glu Glu Asp Arg Ala
    210                 215                 220
Leu Trp Arg Arg Phe Gly Ala Asp Leu Ala Ser Leu Asn Pro Lys Arg
```

```
225                 230                 235                 240
Ile Gly Ala Thr Met Pro Glu Met Ile Ser His Ile His Glu Leu Ile
                245                 250                 255
Asp Glu Arg Arg Ala Ala Leu Arg Asp Asp Leu Leu Ser Gly Leu Ile
            260                 265                 270
Arg Ala Gln Asp Asp Asp Gly Gly Arg Leu Ser Asp Val Glu Met Val
        275                 280                 285
Thr Leu Val Leu Thr Leu Val Leu Ala Gly His Glu Thr Thr Ala His
    290                 295                 300
Leu Ile Ser Asn Gly Thr Leu Ala Leu Leu Thr His Pro Asp Gln Arg
305                 310                 315                 320
Arg Leu Ile Asp Glu Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu
                325                 330                 335
Leu Met Arg Trp Cys Gly Pro Ile Gln Ala Thr Gln Leu Arg Tyr Ala
            340                 345                 350
Leu Glu Asp Thr Glu Val Ala Gly Val Gln Val Arg Gln Gly Glu Ala
        355                 360                 365
Leu Met Phe Ser Leu Val Ala Ala Asn His Asp Pro Arg His Tyr Thr
    370                 375                 380
Gly Pro Glu Arg Leu Asp Leu Thr Arg Gln Pro Ala Gly Arg Ala Glu
385                 390                 395                 400
Asp His Val Gly Phe Gly His Gly Met His Tyr Cys Leu Gly Ala Ser
                405                 410                 415
Leu Ala Arg Gln Glu Ala Glu Val Ala Tyr Gly Lys Leu Leu Thr Arg
            420                 425                 430
Tyr Pro Asp Leu Ala Leu Ala Leu Thr Pro Glu Gln Leu Glu Asp Gln
        435                 440                 445
Glu Arg Leu Arg Gln Pro Gly Thr Trp Arg Leu Arg Arg Leu Pro Leu
    450                 455                 460
Arg Leu His Ala Gln Ser
465                 470
```

<210> 7
<211> 1293
<212> DNA
<213> Streptomyces lydicus

<400> 7

```
atgtcggcat cacccagcaa cacgttcacc gagcacgtcg gcaagcaccc gggcgagccg     60
aacgtgatgg atccggcgct gatcggggat ccgttcgccg gttacggcgc gctgcgcgag    120
cagggcccgg tcgtgcgggg gcggttcatg gacgactccc ccgtgtggtt cgtgacccgc    180
ttcgaggagg tccgcgaggt cctgcgtgac ccgcggttcc ggaacaatcc ggtctccgcg    240
gcgccgggcg cggcccccga ggacaccccg ctgtcccggc tgatggacat gatgggtttc    300
cccgagcacc tgcgcgtcta tctgctcggc tcgatcctca acaacgacgc ccccgaccac    360
acccggctgc gccgcctggt ctcccgggcc ttcaccgcgc ggaagatcac cgatctgcgg    420
ccgcgcgtca cacagatagc cgacgagctg ctggcccggc tgccggagca cgccgaggac    480
ggcgtcgtcg acctgatcca gcacttcgcc tatcccctgc cgatcaccgt catctgcgaa    540
ctggtcggca tccccgagga ggaccgcccg cagtggcgca cctggggcgc cgacctggtc    600
tcgctgcagc cggaccggat gagccggtcc ttcccggcga tgatcgacca catccacgag    660
ctgatcgcgg cgcggcgccg ggcgctcacc gacgatctgc tcagcgagct gatccggacc    720
catgacgacg acggcagccg gctcagcgac gtcgagatgg tcaccatggt cctcaccgtc    780
gtcctggccg gccacgagac caccgcgcac ctcatcggca acggcacggc ggccctgctc    840
acccacccca accagctgcg gctgctcaag gacgacccgg cgctgctgcc gcgcgcggtg    900
```

```
cacgagttga tgcgctggtg cggcccggtg cacatgaccc agctgcgcta cgccgccgag    960
gacgtcgagc tggcgggcgt ccggatccgc acgggggacg ccgtccagct catcctggtg   1020
tcggcgaacc gcgacccgcg ccactacacc gaccccgacc ggctggacct gacccggcac   1080
cctgccggcc acgcggagaa ccatgtgggg ttcggccacg gggcgcacta ctgtctgggc   1140
gccacgctcg ccaagcagga gggcgaggtc gccctcggcg ccctgctcag gcacttcccc   1200
gagctgtcgc tggccgtcgc gccggaggcc ctggagcgca caccggtacc gggcagctgg   1260
cggctgaacg cgctgccgct gcgtctgcgc tga                                1293
```

<210> 8
<211> 430
<212> PRT
<213> Streptomyces lydicus

<400> 8

```
Met Ser Ala Ser Pro Ser Asn Thr Phe Thr Glu His Val Gly Lys His
1               5                  10                  15
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Gly Asp Pro Phe
            20                  25                  30
Ala Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Met Asp Asp Ser Pro Val Trp Phe Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Glu Val Leu Arg Asp Pro Arg Phe Arg Asn Asn Pro Val Ser Ala
65                  70                  75                  80
Ala Pro Gly Ala Ala Pro Glu Asp Thr Pro Leu Ser Arg Leu Met Asp
                85                  90                  95
Met Met Gly Phe Pro Glu His Leu Arg Val Tyr Leu Leu Gly Ser Ile
            100                 105                 110
Leu Asn Asn Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Thr
    130                 135                 140
Gln Ile Ala Asp Glu Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Glu Asp Arg Pro Gln Trp
            180                 185                 190
Arg Thr Trp Gly Ala Asp Leu Val Ser Leu Gln Pro Asp Arg Met Ser
        195                 200                 205
Arg Ser Phe Pro Ala Met Ile Asp His Ile His Glu Leu Ile Ala Ala
    210                 215                 220
Arg Arg Arg Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Ser Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Val Leu Thr Val Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val His Met Thr Gln Leu Arg Tyr Ala Ala Glu
```

```
305                 310                 315                 320
Asp Val Glu Leu Ala Gly Val Arg Ile Arg Thr Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Arg Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Leu Gly Ala Leu Leu Arg His Phe Pro
385                 390                 395                 400
Glu Leu Ser Leu Ala Val Ala Pro Glu Ala Leu Glu Arg Thr Pro Val
                405                 410                 415
Pro Gly Ser Trp Arg Leu Asn Ala Leu Pro Leu Arg Leu Arg
            420                 425                 430
```

<210> 9
<211> 1299
<212> DNA
<213> Streptomyces sp.

<400> 9

```
atgtcagcct tatccagctc tccgttcgcc gagcacatag ggaaacaccc gggcgagccg     60
aacgtgatgg aaccggctct gatcaacgat ccgttcggcg gctacggcgc gctgcgcgag    120
caggggccgg ttgtgcgtgg ccggttcatg gacgactcgc ccgtgtggtt cgtgacccgc    180
ttcgaggagg tccgccaagt cctgcgcgac cagcggttcg tgaacaatcc ggcgtcgccg    240
ctcctgggca gtcaggtcga ggagatgccg atggtcaagc tgctggagca gatgggcctc    300
cccgagcacc ttcgggtcta tctgctcgga tcgatcctca acagtgacgc ccccgatcac    360
acccggcttc gccgcctcgt ctcgcgggcc ttcaccgcac gtaagatcac cggtctgcgg    420
ccgcgcgtcg agcagatcgc cgacgagctg ctggcccggc tccccgagca cgccgaggac    480
ggcgtcgtcg acctcatcca gcacttcgcc tacccgctgc cgatcacggt catctgcgaa    540
ctggtcggca tacccgaagc cgatcgcccg caatggcgcg catggggcgc cgacctcgtg    600
tcactggagc cggacaagct cagcacgtcg ttcccggcga tgatcgacca cacccatgaa    660
ctgatccgcc aacggcgcgg cgcgctcacc gacgatctgc tcagcgagct gatccgtgcc    720
catgacgacg acggcagccg gctcagcgac gtcgagatgg tcaccatggt gttcgctctc    780
gtcttcgccg gtcacgagac caccgcccac ctcataggca acggcacggc ggcgctgctc    840
acccaccccg accagctgcg cctgctcaag gacgacccgg ccctgctccc gcgtgccgtc    900
catgagctga tgcgctggtg cgggccggtg cacatgaccc agttgcgtta cgcctccgag    960
gacatcgacc tcgccggtac gccgatccgg aagggcgacg ccgtccaact catcctggta   1020
tcggcgaact tcgacccccg ccactacagc gaccccgatc gcctcgacct gacccgtcac   1080
cccgcaggcc acgccgagaa ccacgtgggc ttcggccacg ggatgcacta ctgcttgggc   1140
gccgcgctcg ccaggcagga aggcgaagtg gcgttcggca aactgctcgc gcactacccg   1200
gacgtagcgc tgggcgtcga accggaagcc ctggagcggg tgccgatgcc cggcagttgg   1260
cggctgaatt ccttgccgct gcggttggcg aagcgctaa                          1299
```

<210> 10
<211> 432
<212> PRT
<213> Streptomyces sp.

<400> 10

```
Met Ser Ala Leu Ser Ser Ser Pro Phe Ala Glu His Ile Gly Lys His
```

```
1               5                   10                  15
Pro Gly Glu Pro Asn Val Met Glu Pro Ala Leu Ile Asn Asp Pro Phe
            20                  25                  30
Gly Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Met Asp Asp Ser Pro Val Trp Phe Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Gln Val Leu Arg Asp Gln Arg Phe Val Asn Asn Pro Ala Ser Pro
65                  70                  75                  80
Leu Leu Gly Ser Gln Val Glu Glu Met Pro Met Val Lys Leu Leu Glu
                85                  90                  95
Gln Met Gly Leu Pro Glu His Leu Arg Val Tyr Leu Leu Gly Ser Ile
            100                 105                 110
Leu Asn Ser Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Gly Leu Arg Pro Arg Val Glu
    130                 135                 140
Gln Ile Ala Asp Glu Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Ala Asp Arg Pro Gln Trp
            180                 185                 190
Arg Ala Trp Gly Ala Asp Leu Val Ser Leu Glu Pro Asp Lys Leu Ser
        195                 200                 205
Thr Ser Phe Pro Ala Met Ile Asp His Thr His Glu Leu Ile Arg Gln
    210                 215                 220
Arg Arg Gly Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Ala
225                 230                 235                 240
His Asp Asp Asp Gly Ser Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Val Phe Ala Leu Val Phe Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val His Met Thr Gln Leu Arg Tyr Ala Ser Glu
305                 310                 315                 320
Asp Ile Asp Leu Ala Gly Thr Pro Ile Arg Lys Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Phe Asp Pro Arg His Tyr Ser Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Met His Tyr Cys Leu Gly Ala Ala Leu Ala
    370                 375                 380
Arg Gln Glu Gly Glu Val Ala Phe Gly Lys Leu Leu Ala His Tyr Pro
385                 390                 395                 400
Asp Val Ala Leu Gly Val Glu Pro Glu Ala Leu Glu Arg Val Pro Met
                405                 410                 415
Pro Gly Ser Trp Arg Leu Asn Ser Leu Pro Leu Arg Leu Ala Lys Arg
            420                 425                 430
```

<210> 11
<211> 1293
<212> DNA
<213> Streptomyces chattanoogenesis

<400> 11
atgtcggcat cacccagcaa cacgttcacc gagcacgtcg gcaagcaccc gggcgagccg 60
aacgtgatgg atccggcgct gatcggtgat ccgttcgccg gttacggcgc gctgcgcgag 120
cagggcccgg tcgtgcgggg gcggttcatg gacgactccc ccgtgtggtt cgtgacccgc 180
ttcgaggagg tccgcgaggt cctgcgtgac ccgcggttcc ggaacaatcc ggtctccgcg 240
gcgccgggcg cggcccccga ggacaccccg ctgtcccggc tgatggacat gatgggtttc 300
cccgagcacc tgcgcgtcta tctgctcggc tcgatcctca acaacgacgc ccccgaccac 360
acccggctgc gccgcctggt ctcccgggcc ttcaccgcgc ggaagatcac cgatctgcgg 420
ccgcgcgtca cacagatagc cgacgagctg ctggcccggc tgccggagca cgccgaggac 480
ggcgtcgtcg acctgatcca gcacttcgcc tatcccctgc cgatcaccgt catctgcgaa 540
ctggtcggca tccccgagga ggaccgcccg cagtggcgca cctggggcgc cgacctggtc 600
tcgctgcagc cggaccggat gagccggtcc ttcccggcga tgatcgacca catccacgag 660
ctgatcgcgg cgcggcgccg ggcgctcacc gacgacctgc tcagcgagct gatccggacc 720
catgacgacg acggcagcag gctcagcgac gtcgagatgg tcaccatggt cctcaccgtc 780
gtcctggccg gccacgagac caccgcgcac ctcatcggca acggcacggc ggccctgctc 840
acccaccccg accagctgcg gctgctcaag gacgacccgg cactgctgcc gcgcgcggtg 900
cacgagttga tgcgctggtg cggcccggtg cacatgaccc agctgcgcta cgccgccgag 960
gacgtcgagc tggcgggcgt ccggatccgc acgggggacg ccgtccagct catcctggtg 1020
tcggcgaacc gcgacccgcg ccactacacc gaccccgacc gtctggacct gacccggcac 1080
cccgccggtc acgcggagaa ccatgtgggg ttcggccacg gggcgcacta ctgtctgggc 1140
gccacgctcg ccaagcagga gggcgaggtc gccctcggcg ccctgctcag gcacttcccc 1200
gagctgtcgc tggccgtcgc gccggacgcc ctggagcgca caccggtacc gggcagctgg 1260
cggctgaacg cgctgccgct gcgtctgggc tga 1293

<210> 12
<211> 430
<212> PRT
<213> Streptomyces chattanoogenesis

<400> 12
Met Ser Ala Ser Pro Ser Asn Thr Phe Thr Glu His Val Gly Lys His
1 5 10 15
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Gly Asp Pro Phe
20 25 30
Ala Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
35 40 45
Phe Met Asp Asp Ser Pro Val Trp Phe Val Thr Arg Phe Glu Glu Val
50 55 60
Arg Glu Val Leu Arg Asp Pro Arg Phe Arg Asn Asn Pro Val Ser Ala
65 70 75 80
Ala Pro Gly Ala Ala Pro Glu Asp Thr Pro Leu Ser Arg Leu Met Asp
85 90 95
Met Met Gly Phe Pro Glu His Leu Arg Val Tyr Leu Leu Gly Ser Ile
100 105 110
Leu Asn Asn Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
115 120 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Thr

```
    130                 135                 140
Gln Ile Ala Asp Glu Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Glu Asp Arg Pro Gln Trp
            180                 185                 190
Arg Thr Trp Gly Ala Asp Leu Val Ser Leu Gln Pro Asp Arg Met Ser
        195                 200                 205
Arg Ser Phe Pro Ala Met Ile Asp His Ile His Glu Leu Ile Ala Ala
    210                 215                 220
Arg Arg Arg Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Ser Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Val Leu Thr Val Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val His Met Thr Gln Leu Arg Tyr Ala Ala Glu
305                 310                 315                 320
Asp Val Glu Leu Ala Gly Val Arg Ile Arg Thr Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Arg Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Leu Gly Ala Leu Leu Arg His Phe Pro
385                 390                 395                 400
Glu Leu Ser Leu Ala Val Ala Pro Asp Ala Leu Glu Arg Thr Pro Val
                405                 410                 415
Pro Gly Ser Trp Arg Leu Asn Ala Leu Pro Leu Arg Leu Gly
            420                 425                 430
```

<210> 13
<211> 1290
<212> DNA
<213> Streptomyces sp.

<400> 13

```
atgaccgaat tagcggactc ccccttcagc gagcacgtcg gcaaacaccc cggcgagccg     60
aacgtgatgg aaccggccct gctcaccgat ccgttcaccg gctacggcga actgcgcgaa    120
cagggcccgg tggtccgcgg ccggttcgcg gacgacaccc ccgtgtggtt catcacccgc    180
ttcgaggagg cccgcgaggt gctgcgcgac caccggttcg ccaatgcccc cgccttcgcg    240
gcgggaggtg gaagcggtga cacaccctcc aaccggctga tggaaatcat gggcctgccc    300
gagcactacc gggtgtacct cgccaacacc atcctcacca tggacgcccc cgaccacacc    360
cggatccggc gattggtctc ccgggcattc accgcccgta agatcaccga tctgcgaccc    420
cgggtggagg acatcgcgga cgatctgctg aggcggctgc ccgagcacgc cgaggacggc    480
gtcgtcgacc tcatcaagca ctacgcctat ccgctgccca taacggtcat ctgcgaactg    540
```

```
gtgggaattc cggaggaaga ccgactgcag tggcgggatt gggggtccgc gttcgtctcc    600
ctgcaaccgg atcggctcag caaagcgttc ccggcgatga tcgaacacat tcacgcgctg    660
atccgcgaac ggcgcgcggc gctcaccgac gatctgctca gcgaactgat ccgggtccat    720
gacgacgacg gcggccgact cagcgacgtc gaaatggtca cgatggtcct gaccctcgtt    780
ctcgccggtc atgagaccac cgcccatctc atcggcaacg gcactgccgc gcttctcacc    840
caccccgacc agctgcacct gctgaaatcc gatccggagc tgctcccacg cgccgtgcac    900
gagctgatgc gctggtgcgg accggtgcag atgacgcagt tgcggtacgc caccgaggac    960
gtcgaggtgg ccggggtgca ggtcaagcag ggcgaagcgg tgctggccat gctggtcgcg   1020
gcgaaccacg acccccgcca cttcgccgac cccgcccggc tcgacctcac ccgccagccg   1080
gcgggccggg ccgagaacca cgtcggtttc ggccacggca tgcactactg cctgggcgcc   1140
agcctggccc gccaggaggg cgaggtcgcc ttcgggaacc tgctcgcgca ctacccggac   1200
gtgtcgctgg cggtggaacc ggacgccctc cagcgggtcc cgctgccggg caactggcgg   1260
ctggccgcac tgccggtccg gctgcgctga                                    1290
```

<210> 14
<211> 429
<212> PRT
<213> Streptomyces sp.

<400> 14

```
Met Thr Glu Leu Ala Asp Ser Pro Phe Ser Glu His Val Gly Lys His
1               5                  10                  15
Pro Gly Glu Pro Asn Val Met Glu Pro Ala Leu Leu Thr Asp Pro Phe
            20                  25                  30
Thr Gly Tyr Gly Glu Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Ala Asp Asp Thr Pro Val Trp Phe Ile Thr Arg Phe Glu Glu Ala
    50                  55                  60
Arg Glu Val Leu Arg Asp His Arg Phe Ala Asn Ala Pro Ala Phe Ala
65                  70                  75                  80
Ala Gly Gly Gly Ser Gly Asp Thr Pro Ser Asn Arg Leu Met Glu Ile
                85                  90                  95
Met Gly Leu Pro Glu His Tyr Arg Val Tyr Leu Ala Asn Thr Ile Leu
            100                 105                 110
Thr Met Asp Ala Pro Asp His Thr Arg Ile Arg Arg Leu Val Ser Arg
        115                 120                 125
Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu Asp
    130                 135                 140
Ile Ala Asp Asp Leu Leu Arg Arg Leu Pro Glu His Ala Glu Asp Gly
145                 150                 155                 160
Val Val Asp Leu Ile Lys His Tyr Ala Tyr Pro Leu Pro Ile Thr Val
                165                 170                 175
Ile Cys Glu Leu Val Gly Ile Pro Glu Glu Asp Arg Leu Gln Trp Arg
            180                 185                 190
Asp Trp Gly Ser Ala Phe Val Ser Leu Gln Pro Asp Arg Leu Ser Lys
        195                 200                 205
Ala Phe Pro Ala Met Ile Glu His Ile His Ala Leu Ile Arg Glu Arg
    210                 215                 220
Arg Ala Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Val His
225                 230                 235                 240
Asp Asp Asp Gly Gly Arg Leu Ser Asp Val Glu Met Val Thr Met Val
                245                 250                 255
Leu Thr Leu Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile Gly
```

```
            260                 265                 270
Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu His Leu Leu
        275                 280                 285
Lys Ser Asp Pro Glu Leu Leu Pro Arg Ala Val His Glu Leu Met Arg
    290                 295                 300
Trp Cys Gly Pro Val Gln Met Thr Gln Leu Arg Tyr Ala Thr Glu Asp
305                 310                 315                 320
Val Glu Val Ala Gly Val Gln Val Lys Gln Gly Glu Ala Val Leu Ala
                325                 330                 335
Met Leu Val Ala Ala Asn His Asp Pro Arg His Phe Ala Asp Pro Ala
            340                 345                 350
Arg Leu Asp Leu Thr Arg Gln Pro Ala Gly Arg Ala Glu Asn His Val
        355                 360                 365
Gly Phe Gly His Gly Met His Tyr Cys Leu Gly Ala Ser Leu Ala Arg
    370                 375                 380
Gln Glu Gly Glu Val Ala Phe Gly Asn Leu Leu Ala His Tyr Pro Asp
385                 390                 395                 400
Val Ser Leu Ala Val Glu Pro Asp Ala Leu Gln Arg Val Pro Leu Pro
                405                 410                 415
Gly Asn Trp Arg Leu Ala Ala Leu Pro Val Arg Leu Arg
            420                 425
```

<210> 15
<211> 1428
<212> DNA
<213> *Streptomyces albofaciens*

<400> 15

```
atgaccacat cgcccaccga gtcccgggcg gccaccccgc ccgactccac cgcctccccc     60
tcgaccgctg ccgccccggc caccacccct tcggccgccg cctctccgga caccacctct    120
cccgccacca ccgaccgcac cacgctcccc tcctacgtcg gcctccaccc gggcgagccg    180
aacctgatgg aaccggagct gctggacaac ccgtacaccg gctacggcac gctgcgcgag    240
caggcgccgc tcgtccgcgc ccggttcatc gacgactcgc ccatctggct ggtgacccgc    300
ttcgacgtgg tgcgcgaggt gatgcgcgac cagcggttcg tcaacaaccc gaccctggtg    360
cccggcatcg gtgcggacca ggacccgcgc gcccggctga tcgagctgtt cggcatcccc    420
gaggacctgg ccccgtacct caccgacacc atcctcacca gcgacccgcc ggaccacacc    480
cggctgcgcc gcctggtctc ccgtgccttc accgcacgcc gtatccagga cctgcggccg    540
cgcgtcgagc ggatcaccga cgagctgctg gcgcggctgc cggaccatgc cgaggacggc    600
gtcgtcgacc tcgtcgagca cttcgcctac ccgctgccca tcacggtcat ctgcgaactg    660
gtcggcatcg acgaggagga ccgggcgctg tggcggcggt tcggcgccga cctcgcctcg    720
ctgaacccca agcgcatcgg cgccaccatg ccggagatga tcgcgcacat ccacgaggtg    780
atcgacgagc ggcgtgcgga cctgcgggac gacctgctca gcgggctcat ccgggcgcag    840
gacgacgacg gcggccggct gagcgacgtc gagatggtca cgctggtgct gaccctggtg    900
ctggccggtc acgagaccac cgcccacctc atcagcaacg gcaccctcgc cctgctcacc    960
caccccgacc agcggcggct gatcgacgag gacccggcgc tgctgccgcg cgcggtccac   1020
gagctgatgc gctggtgcgg gccgatccag gccacccagc tgcggtacgc catggaggac   1080
accgaggtgg ccggtgtcca ggtccgccag ggcgaggccc tgatgttcag cctcgtcgcg   1140
gccaaccacg acccgcgcca ctacaccggc ccggagcggc tcgacctgac gcggcagccg   1200
gccggccgcg ccgaggacca cgtcggcttc gggcacggga tgcactactg cctgggtgcc   1260
tcactggccc ggcaggaggc cgaggtggcg tacggcaagc tgctcacccg ctacccggac   1320
ctggcgctcg cgctcacccc ggaacagctg gaggaccagg aacgcctgcg gcagcccggc   1380
acctggcgcc tgcgacggct gccgctgagg ttgcacgcgg agagctga                1428
```

<210> 16
<211> 475
<212> PRT
<213> Streptomyces albofaciens

<400> 16

```
Met Thr Thr Ser Pro Thr Glu Ser Arg Ala Ala Thr Pro Pro Asp Ser
1               5                  10                  15
Thr Ala Ser Pro Ser Thr Ala Ala Ala Pro Ala Thr Thr Pro Ser Ala
            20                  25                  30
Ala Ala Ser Pro Asp Thr Thr Ser Pro Ala Thr Thr Asp Arg Thr Thr
        35                  40                  45
Leu Pro Ser Tyr Val Gly Leu His Pro Gly Glu Pro Asn Leu Met Glu
    50                  55                  60
Pro Glu Leu Leu Asp Asn Pro Tyr Thr Gly Tyr Gly Thr Leu Arg Glu
65                  70                  75                  80
Gln Ala Pro Leu Val Arg Ala Arg Phe Ile Asp Asp Ser Pro Ile Trp
                85                  90                  95
Leu Val Thr Arg Phe Asp Val Val Arg Glu Val Met Arg Asp Gln Arg
            100                 105                 110
Phe Val Asn Asn Pro Thr Leu Val Pro Gly Ile Gly Ala Asp Gln Asp
        115                 120                 125
Pro Arg Ala Arg Leu Ile Glu Leu Phe Gly Ile Pro Glu Asp Leu Ala
    130                 135                 140
Pro Tyr Leu Thr Asp Thr Ile Leu Thr Ser Asp Pro Pro Asp His Thr
145                 150                 155                 160
Arg Leu Arg Arg Leu Val Ser Arg Ala Phe Thr Ala Arg Arg Ile Gln
                165                 170                 175
Asp Leu Arg Pro Arg Val Glu Arg Ile Thr Asp Glu Leu Leu Ala Arg
            180                 185                 190
Leu Pro Asp His Ala Glu Asp Gly Val Val Asp Leu Val Glu His Phe
        195                 200                 205
Ala Tyr Pro Leu Pro Ile Thr Val Ile Cys Glu Leu Val Gly Ile Asp
    210                 215                 220
Glu Glu Asp Arg Ala Leu Trp Arg Arg Phe Gly Ala Asp Leu Ala Ser
225                 230                 235                 240
Leu Asn Pro Lys Arg Ile Gly Ala Thr Met Pro Glu Met Ile Ala His
                245                 250                 255
Ile His Glu Val Ile Asp Glu Arg Arg Ala Asp Leu Arg Asp Asp Leu
            260                 265                 270
Leu Ser Gly Leu Ile Arg Ala Gln Asp Asp Asp Gly Gly Arg Leu Ser
        275                 280                 285
Asp Val Glu Met Val Thr Leu Val Leu Thr Leu Val Leu Ala Gly His
    290                 295                 300
Glu Thr Thr Ala His Leu Ile Ser Asn Gly Thr Leu Ala Leu Leu Thr
305                 310                 315                 320
His Pro Asp Gln Arg Arg Leu Ile Asp Glu Asp Pro Ala Leu Leu Pro
                325                 330                 335
Arg Ala Val His Glu Leu Met Arg Trp Cys Gly Pro Ile Gln Ala Thr
            340                 345                 350
Gln Leu Arg Tyr Ala Met Glu Asp Thr Glu Val Ala Gly Val Gln Val
        355                 360                 365
Arg Gln Gly Glu Ala Leu Met Phe Ser Leu Val Ala Ala Asn His Asp
```

```
    370                 375                 380
Pro Arg His Tyr Thr Gly Pro Glu Arg Leu Asp Leu Thr Arg Gln Pro
385                 390                 395                 400
Ala Gly Arg Ala Glu Asp His Val Gly Phe Gly His Gly Met His Tyr
                405                 410                 415
Cys Leu Gly Ala Ser Leu Ala Arg Gln Glu Ala Glu Val Ala Tyr Gly
            420                 425                 430
Lys Leu Leu Thr Arg Tyr Pro Asp Leu Ala Leu Ala Leu Thr Pro Glu
        435                 440                 445
Gln Leu Glu Asp Gln Glu Arg Leu Arg Gln Pro Gly Thr Trp Arg Leu
    450                 455                 460
Arg Arg Leu Pro Leu Arg Leu His Ala Glu Ser
465                 470                 475
```

<210> 17
<211> 1293
<212> DNA
<213> Streptomyces platensis

<400> 17

```
atgtcggcat tacccacctc accgttcgct gcacacgtcg ggaaacaccc gggcgagccg     60
aatgtgatgg acccggcact gatcaccgac ccgttcaccg gctacggcgc gctgcgcgag    120
cagggcccgg tcgtccgcgg ccgcttcgtg gacgactcac ccgtctggct ggtgacgcga    180
ttcgaggagg tccgccaagt cctgcgcgac cagcggttcg tgaacaaccc ggcggcgccc    240
tccctgggcc acgcggccga ggacaacccg ctcaccaggc tgatggacat gctgggcctc    300

cccgagcacc tccgccccta cctcctcgga tcgattctca attacgacgc ccccgaccac    360
acccggctgc gccgcctggt gtcgcgggcc ttcaccgccc gcaagatcac cgacctgcgg    420
ccgcgggtcg agcagatcgc cgacgccctg ctggcccggc tgcccgagca cgccgaggac    480
ggcgtcgtcg atctcatccg gcacttcgcc tacccgctgc cgatcaccgt catctgcgaa    540
ctggtcggca tacccgaagc ggaccgcccg cagtggcgga cgtggggcgc cgacctcgtc    600
tcgatggagc cggaccggct caccgcctcg ttcccgccga tgatcgagca catccaccgg    660
atggtccggg agcggcgcgg cgcgctcacc ggcgatctgc tcagcgagct gatccgtgcc    720
catgacgacg acggcggccg gctcagcgac gtcgagatgg tcaccttgat cctcacgctc    780
gtcctcgccg gtcacgagac caccgctcac ctcatcagca acggcacggc ggcgctgctc    840
acccaccccg accaactgcg cctgctccag gacgacccgg ccctgctccc ccgtgccgtc    900
cacgagctga tgcgctggtg cgggccggtg cagatgaccc agctgcgtta cgccgccgcc    960
gacgtcgacc tggccggcac cacgatccac cggggcgacg ccgtccaact catcctggtg   1020
tcggcgaact tcgaccccg ccactacacc gaccccgacc gcctcgatct gacccgccac   1080
cccgcgggac atgcggagaa ccatgtgggt ttcggccatg gggcgcacta ctgcctgggc   1140
gccacactcg ccaagcagga gggcgaagtc gccttcggca aactgctcgc gcactacccg   1200
gagatggcgt tgggcgtcgc accggagcgc ctggagcgga cgcccctgcc gggcaactgg   1260
cggctgaacg cgctgccggt gcggttgggg tga                                1293
```

<210> 18
<211> 430
<212> PRT
<213> Streptomyces platensis

<400> 18

```
Met Ser Ala Leu Pro Thr Ser Pro Phe Ala Ala His Val Gly Lys His
1               5                   10                  15
```

```
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Thr Asp Pro Phe
            20                  25                  30
Thr Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Val Asp Asp Ser Pro Val Trp Leu Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Gln Val Leu Arg Asp Gln Arg Phe Val Asn Asn Pro Ala Ala Pro
65                  70                  75                  80
Ser Leu Gly His Ala Ala Glu Asp Asn Pro Leu Thr Arg Leu Met Asp
                85                  90                  95
Met Leu Gly Leu Pro Glu His Leu Arg Pro Tyr Leu Leu Gly Ser Ile
            100                 105                 110
Leu Asn Tyr Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu
    130                 135                 140
Gln Ile Ala Asp Ala Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Arg His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Ala Asp Arg Pro Gln Trp
            180                 185                 190
Arg Thr Trp Gly Ala Asp Leu Val Ser Met Glu Pro Asp Arg Leu Thr
        195                 200                 205
Ala Ser Phe Pro Pro Met Ile Glu His Ile His Arg Met Val Arg Glu
    210                 215                 220
Arg Arg Gly Ala Leu Thr Gly Asp Leu Leu Ser Glu Leu Ile Arg Ala
225                 230                 235                 240
His Asp Asp Asp Gly Gly Arg Leu Ser Asp Val Glu Met Val Thr Leu
                245                 250                 255
Ile Leu Thr Leu Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Ser Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Gln Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val Gln Met Thr Gln Leu Arg Tyr Ala Ala Ala
305                 310                 315                 320
Asp Val Asp Leu Ala Gly Thr Thr Ile His Arg Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Phe Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Phe Gly Lys Leu Leu Ala His Tyr Pro
385                 390                 395                 400
Glu Met Ala Leu Gly Val Ala Pro Glu Arg Leu Glu Arg Thr Pro Leu
                405                 410                 415
Pro Gly Asn Trp Arg Leu Asn Ala Leu Pro Val Arg Leu Gly
            420                 425                 430
```

<210> 19
<211> 1293
<212> DNA
<213> Streptomyces kasugaensis

<400> 19
atgtcggcat cacccagcaa cacgttcacc gagcacgtcg gcaagcaccc gggcgagccg 60
aacgtgatgg atccggcgct gatcggggat ccgttcgccg gttacggcgc gctgcgcgag 120
cagggcccgg tcgtgcgggg gcggttcatg gacgactccc ccgtgtggtt cgtgacccgc 180
ttcgaggagg tccgcgaggt cctgcgtgac ccgcggttcc ggaacaatcc ggtctccgcg 240
gcgccgggcg cggcccccga ggacaccccg ctgtcccggc tgatggacat gatgggtttc 300
cccgagcacc tgcgcgtcta tctgctcggc tcgatcctca acaacgacgc ccccgaccac 360
acccggctgc gccgcctggt ctcccgggcc ttcaccgcgc ggaagatcac cgatctgcgg 420
ccgcgcgtca cacagatagc cgacgagctg ctggcccggc tgccggagca cgccgaggac 480
ggcgtcgtcg acctgatcca gcacttcgcc tatcccctgc cgatcaccgt catctgcgaa 540
ctggtcggca tccccgagga ggaccgcccg cagtggcgca cctggggcgc cgacctggtc 600
tcgctgcagc cggaccggat gagccggtcc ttcccggcga tgatcgacca catccacgag 660
ctgatcgcgg cgcggcgccg ggcgctcacc gacgatctgc tcagcgagct gatccggacc 720
catgacgacg acggcagccg gctcagcgac gtcgagatgg tcaccatggt cctcaccgtc 780
gtcctggccg gccacgagac caccgcgcac ctcatcggca acggcacggc ggccctgctc 840
acccaccccg accagctgcg gctgctcaag gacgacccgg cgctgctgcc gcgcgcggtg 900
cacgagttga tgcgctggtg cggcccggtg cacatgaccc agctgcgcta cgccgccgag 960
gacgtcgagc tggcgggcgt ccggatccgc acgggggacg ccgtccagct catcctggtg 1020
tcggcgaacc gcgacccgcg ccactacacc gaccccgacc ggctggacct gacccggcac 1080
cctgccggcc acgcggagaa ccatgtgggg ttcggccacg gggcgcacta ctgtctgggc 1140
gccacgctcg ccaagcagga gggcgaggtc gccctcggcg ccctgctcag gcacttcccc 1200
gagctgtcgc tggccgtcgc gccggaggcc ctggagcgca caccggtacc gggcagctgg 1260
cggctgaacg cgctgccgct gcgtctgcgc tga 1293

<210> 20
<211> 430
<212> PRT
<213> Streptomyces kasugaensis

<400> 20
Met Ser Ala Ser Pro Ser Asn Thr Phe Thr Glu His Val Gly Lys His
1 5 10 15
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Gly Asp Pro Phe
20 25 30
Ala Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
35 40 45
Phe Met Asp Asp Ser Pro Val Trp Phe Val Thr Arg Phe Glu Glu Val
50 55 60
Arg Glu Val Leu Arg Asp Pro Arg Phe Arg Asn Asn Pro Val Ser Ala
65 70 75 80
Ala Pro Gly Ala Ala Pro Glu Asp Thr Pro Leu Ser Arg Leu Met Asp
85 90 95
Met Met Gly Phe Pro Glu His Leu Arg Val Tyr Leu Leu Gly Ser Ile
100 105 110
Leu Asn Asn Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
115 120 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Thr
130 135 140

```
Gln Ile Ala Asp Glu Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Glu Asp Arg Pro Gln Trp
            180                 185                 190
Arg Thr Trp Gly Ala Asp Leu Val Ser Leu Gln Pro Asp Arg Met Ser
        195                 200                 205
Arg Ser Phe Pro Ala Met Ile Asp His Ile His Glu Leu Ile Ala Ala
    210                 215                 220
Arg Arg Arg Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Ser Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Val Leu Thr Val Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val His Met Thr Gln Leu Arg Tyr Ala Ala Glu
305                 310                 315                 320
Asp Val Glu Leu Ala Gly Val Arg Ile Arg Thr Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Arg Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Leu Gly Ala Leu Leu Arg His Phe Pro
385                 390                 395                 400
Glu Leu Ser Leu Ala Val Ala Pro Glu Ala Leu Glu Arg Thr Pro Val
                405                 410                 415
Pro Gly Ser Trp Arg Leu Asn Ala Leu Pro Leu Arg Leu Arg
            420                 425                 430
```

<210> 21
<211> 1428
<212> DNA
<213> Streptomyces rimosus

<400> 21

```
atgaccacat cgcccaccga gtcccgggcg gccaccccga ccggctccac cgcctccccc     60
tcgaccgctt ccgccccggc caccaccccc tcggccgcca cctcttcgga caccacctat    120
cccgccacca ccgaccgcac cacgctcccc tcctacgtcg gcctccaccc gggcgagccg    180
aacctgatgg aaccggagct gctggacaac ccgtacaccg gctacggcac gctgcgcgag    240
caggccccgc tcgtccgtgc ccggttcatc gacgactcgc ccatctggct ggtgacccgc    300
ttcgacgtgg tgcgcgaggt gatgcgcgac cagcggttcg tcaacaaccc gaccctggtg    360
cccggcatcg gtgcggacaa ggacccgcgc gcccggctga tcgagctgtt cggcatcccc    420
gaggacctga ccccgtacct cgccgacacc atcctcacca gcgacccgcc ggaccacacc    480
cggctgcgcc gcctggtctc ccgtgccttc accgcgcgcc gcatccagga cctgcggccg    540
cgcgtcgagc agatcaccga cgcgctgctg gagcgactgc cggaccatgc cgaggacggc    600
```

```
gtcgtcgacc tcgtcgagca cttcgcctac ccgctgccca tcacggtcat ctgcgagctg    660
gtcggcatcg acgaggagga ccggacgctg tggcggcggt tcggcgccga cctcgcctca    720
ctgaacccca agcgcatcgg cgccaccatg ccggagatga tcgcgcacat ccacgaggtg    780
atcgacgagc ggcgcgcggc cctgcgggac gacctgctca gcgggctcat ccgggcgcag    840
gacgacgacg gcggccggct gagcgacgtc gagatggtca ccctggtcct gaccctggtg    900
ctggccggtc acgagaccac cgcccacctc atcagcaacg gcaccctcgc cctgctcacc    960
caccccgacc agcggcggct gatcgacgag gacccggcac tgctgccgcg cgcggtccac   1020
gagctgatgc gctggtgcgg gccgatccag gccacccagc tgcggtacgc catggaggac   1080
accgaggtcg ccggtgtcca ggtccgccag ggcgaggccc tgatgttcag cctcgtcgcg   1140
gccaaccacg acccgcgcca ctacaccggg ccggagcggc tcgacctgac gcggcagccg   1200
gccggccgcg ccgaggacca cgtcggcttc gggcacggga tgcactactg cctgggtgcc   1260
tcactcgccc ggcaggaggc cgaggtggcc tacgggaagc tgctcacccg ctacccggac   1320
ctggagctcg ctctcacacc ggaacagctg gaggaccagg aacgcctgcg gcagcccggc   1380
acctggcgcc tgcggcggct gccgctgaag ctgcacgcgc ggagctga                1428
```

<210> 22
<211> 475
<212> PRT
<213> Streptomyces rimosus

<400> 22

```
Met Thr Thr Ser Pro Thr Glu Ser Arg Ala Ala Thr Pro Thr Gly Ser
1               5                  10                  15
Thr Ala Ser Pro Ser Thr Ala Ser Ala Pro Ala Thr Thr Pro Ser Ala
            20                  25                  30
Ala Thr Ser Ser Asp Thr Thr Tyr Pro Ala Thr Thr Asp Arg Thr Thr
        35                  40                  45
Leu Pro Ser Tyr Val Gly Leu His Pro Gly Glu Pro Asn Leu Met Glu
    50                  55                  60
Pro Glu Leu Leu Asp Asn Pro Tyr Thr Gly Tyr Gly Thr Leu Arg Glu
65                  70                  75                  80
Gln Ala Pro Leu Val Arg Ala Arg Phe Ile Asp Asp Ser Pro Ile Trp
                85                  90                  95
Leu Val Thr Arg Phe Asp Val Val Arg Glu Val Met Arg Asp Gln Arg
            100                 105                 110
Phe Val Asn Asn Pro Thr Leu Val Pro Gly Ile Gly Ala Asp Lys Asp
        115                 120                 125
Pro Arg Ala Arg Leu Ile Glu Leu Phe Gly Ile Pro Glu Asp Leu Thr
    130                 135                 140
Pro Tyr Leu Ala Asp Thr Ile Leu Thr Ser Asp Pro Pro Asp His Thr
145                 150                 155                 160
Arg Leu Arg Arg Leu Val Ser Arg Ala Phe Thr Ala Arg Arg Ile Gln
                165                 170                 175
Asp Leu Arg Pro Arg Val Glu Gln Ile Thr Asp Ala Leu Leu Glu Arg
            180                 185                 190
Leu Pro Asp His Ala Glu Asp Gly Val Val Asp Leu Val Glu His Phe
        195                 200                 205
Ala Tyr Pro Leu Pro Ile Thr Val Ile Cys Glu Leu Val Gly Ile Asp
    210                 215                 220
Glu Glu Asp Arg Thr Leu Trp Arg Arg Phe Gly Ala Asp Leu Ala Ser
225                 230                 235                 240
Leu Asn Pro Lys Arg Ile Gly Ala Thr Met Pro Glu Met Ile Ala His
                245                 250                 255
```

```
Ile His Glu Val Ile Asp Glu Arg Arg Ala Ala Leu Arg Asp Asp Leu
            260                 265                 270
Leu Ser Gly Leu Ile Arg Ala Gln Asp Asp Asp Gly Gly Arg Leu Ser
        275                 280                 285
Asp Val Glu Met Val Thr Leu Val Leu Thr Leu Val Leu Ala Gly His
    290                 295                 300
Glu Thr Thr Ala His Leu Ile Ser Asn Gly Thr Leu Ala Leu Leu Thr
305                 310                 315                 320
His Pro Asp Gln Arg Arg Leu Ile Asp Glu Asp Pro Ala Leu Leu Pro
                325                 330                 335
Arg Ala Val His Glu Leu Met Arg Trp Cys Gly Pro Ile Gln Ala Thr
            340                 345                 350
Gln Leu Arg Tyr Ala Met Glu Asp Thr Glu Val Ala Gly Val Gln Val
        355                 360                 365
Arg Gln Gly Glu Ala Leu Met Phe Ser Leu Val Ala Ala Asn His Asp
    370                 375                 380
Pro Arg His Tyr Thr Gly Pro Glu Arg Leu Asp Leu Thr Arg Gln Pro
385                 390                 395                 400
Ala Gly Arg Ala Glu Asp His Val Gly Phe Gly His Gly Met His Tyr
                405                 410                 415
Cys Leu Gly Ala Ser Leu Ala Arg Gln Glu Ala Glu Val Ala Tyr Gly
            420                 425                 430
Lys Leu Leu Thr Arg Tyr Pro Asp Leu Glu Leu Ala Leu Thr Pro Glu
        435                 440                 445
Gln Leu Glu Asp Gln Glu Arg Leu Arg Gln Pro Gly Thr Trp Arg Leu
    450                 455                 460
Arg Arg Leu Pro Leu Lys Leu His Ala Arg Ser
465                 470                 475
```

<210> 23
<211> 1293
<212> DNA
<213> Streptomyces tubercidicus

<400> 23

```
atgtcggcat tatccaactc cccgctcgcc gcacatgtcg ggaaacaccc tggcgagccg     60
aatgtgatgg acccggcgct gatcaccgac ccgttcggcg gctacggcgc actgcgcgag    120
caaggcccgg tcgtacgggg ccggttcatg gacgactcgc ccgtctggct ggtgacgcgc    180
ttcgaagagg tccgccaagt cctgcgcgat cagcggttcg tgaacaaccc ggccgcaccg    240
tccctgggac gctcgatcga cgaaagcccc gcggtcagac ttttggaaat gttggggttg    300
cccgaccatt tccggccgta tctgctcggg tcgatcctca actacgacgc acccgaccac    360
acccggctcc gccgactggt ctcgcgcgcc ttcacggcac gcaagatcac cgacctgcgg    420
ccgcgggtcg agcagatcac cgacgacctg ctgacccggc ttcccgagca cgccgaggac    480
ggtgtggtcg acctcatcca gcacttcgcc taccccctgc cgatcaccgt gatctgcgaa    540

ctggtcggca tcgccgaagc ggaccgcccg caatggcgga agtggggagc cgatctcgtc    600
tcgctggagc cggggcggct gagcaccgcg ttcccggcga tggtcgagca catccatgag    660
ctgatccgcg agcggcgcgg cgcgctcacc gacgatctgc tcagcgagct gatccgcacc    720
catgacgacg acggcggccg gctcagcgac atcgagatgg tcaccatgat cctcacgatc    780
gtcctggccg gccacgagac caccgcccac ctcataggca acggcacggc ggcgctgctc    840
acccaccccg accagctgcg cctactcaag gacgatccgg cgctgctgcc gcgcgccgtc    900
cacgagctga tgcgctggtg cgggccggtg cacatgaccc agctgcggtt cgcgtccgag    960
gacgtcgagg tcgccgggac accgatccac aagggcgacg ccgtacaact catcctggta   1020
```

```
tcggcgaact tcgacccccg ccactacacc gaccccgacc gtctcgacct gacccgccac   1080
cccgccggcc acgccgagaa ccatgtgggc ttcggccacg gaatgcacta ctgcctgggt   1140
gccaccctcg ccaaacagga aggcgaagtc gccttctccc gcctcttcac gcactacccg   1200
gaactgtccc tgggcgtcgc ggcggaccag ctggcgcgga cacaggtacc cggcagctgg   1260
cggctggaca ccctgccgct gcgactgggg tga                                1293
```

<210> 24
<211> 430
<212> PRT
<213> Streptomyces tubercidicus

<400> 24

```
Met Ser Ala Leu Ser Asn Ser Pro Leu Ala Ala His Val Gly Lys His
1               5                  10                  15
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Thr Asp Pro Phe
            20                  25                  30
Gly Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Met Asp Asp Ser Pro Val Trp Leu Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Gln Val Leu Arg Asp Gln Arg Phe Val Asn Asn Pro Ala Ala Pro
65                  70                  75                  80
Ser Leu Gly Arg Ser Ile Asp Glu Ser Pro Ala Val Arg Leu Leu Glu
                85                  90                  95
Met Leu Gly Leu Pro Asp His Phe Arg Pro Tyr Leu Leu Gly Ser Ile
            100                 105                 110
Leu Asn Tyr Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu
    130                 135                 140
Gln Ile Thr Asp Asp Leu Leu Thr Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Ala Glu Ala Asp Arg Pro Gln Trp
            180                 185                 190
Arg Lys Trp Gly Ala Asp Leu Val Ser Leu Glu Pro Gly Arg Leu Ser
        195                 200                 205
Thr Ala Phe Pro Ala Met Val Glu His Ile His Glu Leu Ile Arg Glu
    210                 215                 220
Arg Arg Gly Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Gly Arg Leu Ser Asp Ile Glu Met Val Thr Met
                245                 250                 255
Ile Leu Thr Ile Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val His Met Thr Gln Leu Arg Phe Ala Ser Glu
305                 310                 315                 320
Asp Val Glu Val Ala Gly Thr Pro Ile His Lys Gly Asp Ala Val Gln
```

```
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Phe Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Met His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Phe Ser Arg Leu Phe Thr His Tyr Pro
385                 390                 395                 400
Glu Leu Ser Leu Gly Val Ala Ala Asp Gln Leu Ala Arg Thr Gln Val
                405                 410                 415
Pro Gly Ser Trp Arg Leu Asp Thr Leu Pro Leu Arg Leu Gly
            420                 425                 430
```

<210> 25
<211> 1293
<212> DNA
<213> Streptomyces platensis

<400> 25

```
atgtcggcat tatccagctc accgttcgcc gcgcatgtcg ggaaacaccc gggcgagccg     60
aatgtgatgg acccggcgct gatcgccgat ccgttcggtg gttatggcgc actgcgtgag    120
caagggccgg tcgtacgggg ccggttcatg gacgactcac ccgtctggct cgtgacgcgc    180
ttcgaggaag tccgccaagt cctgcgcgac cagcggttcc tgaacgatcc gacggccccc    240
tccctggggc gctcattcga cgacagcccc acggccaggc tgctggagat gatgggactg    300
cccgagcatt tccggccgta tctgctcggt tcgattctga acaacgacgc ccccgaccac    360
acccggctgc gccgtctggt gtcgcgcgcc ttcacggcac gcaagatcac cgacctgcgg    420
ccgcgggtcg agcagatcgc cgacgagctg ctgacccggc ttcccgagta cgccgaggac    480
ggcgtggtcg acctcatcaa gcacttcgcc tacccctgc cgatcgccgt catctgcgaa    540
ctggtcggca tagccgaagc ggatcgtccg cagtggcgga agtggggtgc cgacctcgtc    600
tcgctgcagc cggaccggct cagcacctcg ttcccggcga tgatcgagca catccatgag    660
ctgatccgcg agcggcgcgg ggcgctcacg gacgatctgc tcagcgagct gatccgtgcc    720
catgacgacg acggcggccg gctcagcgac gtcgagatgg tcaccatgat cctcacggtg    780
gtgctcgccg gccacgagac caccgcgcac ctcataggca acggcacggc ggcgctgctc    840
acccaccccg accagctgcg gctgctcagg gacgacccgg ctctgtttcc ccgtgccgtc    900
cacgagctgt tgcgctggtg cgggccggtc cacatgaccc agatgcggtt tgcgtccgag    960
gatgtcgaca tcgccgggac gaagatccgt aagggcgacg ccgtacaact gatcctggta   1020
tcggccaact tcgacccccg ccactacacc gaccccgaac gtctcgacct gacccgtcac   1080
cccgccggcc acgccgagaa ccatgtgggc ttcggccacg ggatgcacta ctgcctgggc   1140
gccaccctcg ccaaacagga gggcgaagtc gcgttcgaga agctcttcgc gcactacccg   1200
gaggtgtcgc tgggcgtcgc accggaacaa ctggaaagga caccactgcc cggcagctgg   1260
cggctcgatt ccctgccgct gcggttgcgg taa                                1293
```

<210> 26
<211> 430
<212> PRT
<213> Streptomyces platensis

<400> 26

```
Met Ser Ala Leu Ser Ser Ser Pro Phe Ala Ala His Val Gly Lys His
1               5                   10                  15
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Ala Asp Pro Phe
```

```
            20                  25                  30
Gly Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Met Asp Asp Ser Pro Val Trp Leu Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Gln Val Leu Arg Asp Gln Arg Phe Leu Asn Asp Pro Thr Ala Pro
65                  70                  75                  80
Ser Leu Gly Arg Ser Phe Asp Asp Ser Pro Thr Ala Arg Leu Leu Glu
                85                  90                  95
Met Met Gly Leu Pro Glu His Phe Arg Pro Tyr Leu Leu Gly Ser Ile
            100                 105                 110
Leu Asn Asn Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu
    130                 135                 140
Gln Ile Ala Asp Glu Leu Leu Thr Arg Leu Pro Glu Tyr Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Lys His Phe Ala Tyr Pro Leu Pro Ile Ala
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Ala Glu Ala Asp Arg Pro Gln Trp
            180                 185                 190
Arg Lys Trp Gly Ala Asp Leu Val Ser Leu Gln Pro Asp Arg Leu Ser
        195                 200                 205
Thr Ser Phe Pro Ala Met Ile Glu His Ile His Glu Leu Ile Arg Glu
    210                 215                 220
Arg Arg Gly Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Ala
225                 230                 235                 240
His Asp Asp Asp Gly Gly Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Ile Leu Thr Val Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Arg Asp Asp Pro Ala Leu Phe Pro Arg Ala Val His Glu Leu Leu
    290                 295                 300
Arg Trp Cys Gly Pro Val His Met Thr Gln Met Arg Phe Ala Ser Glu
305                 310                 315                 320
Asp Val Asp Ile Ala Gly Thr Lys Ile Arg Lys Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Phe Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Glu Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Met His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Phe Glu Lys Leu Phe Ala His Tyr Pro
385                 390                 395                 400
Glu Val Ser Leu Gly Val Ala Pro Glu Gln Leu Glu Arg Thr Pro Leu
                405                 410                 415
Pro Gly Ser Trp Arg Leu Asp Ser Leu Pro Leu Arg Leu Arg
            420                 425                 430
```

<210> 27

<211> 1293
<212> DNA
<213> Streptomyces platensis

<400> 27
atgtcggcat tatccagctc tccgttcgct gcgcatgtcg ggaaacaccc gggtgagccg 60
aatgtgatgg agccggcgct gctcaccgac ccgttcgcgg gctacggcgc gctgcgtgag 120
caggccccgg tcgtacgggg ccggttcgtg gacgactcac cggtctggtt cgtgacgcgc 180
ttcgaggagg tccgccaagt cctgcgcgac cagcggttcg tgaacaatcc ggccgcgccg 240
cccctggccc catcggccga ggagaacccg ctgaccaggc tgatggacat gctgggcctc 300
cccgagcacc tccgcgtcta catgctcggg tcgattctca actacgacgc ccccgaccac 360
acccggctgc gccgtctggt gtcgcgcgcg ttcacggcgc ggaagatcac cgatctgcga 420
ccgcgtgtcg agcagatcgc cgacgagctg ctggcccgcc tccccgagta cgccgaggac 480
ggcgtcgtcg acctcatcca gcatttcgcc tacccgctgc cgatcaccgt catctgcgag 540
ctggtcggca tacccgaagc ggaccgcccg cagtggcgga agtggggcgc cgacctcatc 600
tcgatggacc cggaccggct cggcgcaacg ttcccggcga tgatcgagca catccatgag 660
atggtccggg agcggcgcgc ggcgctcacc gatgatctgc tcagcgagct gatccgtacc 720
catgacgacg atggcggccg gctcagcgac gtcgagatgg tcaccatgat cctcacgctc 780
gtcctcgccg gtcacgagac caccgcccac ctcatcagca acggcacggc ggcgctgctc 840
acccaccccg accagctgcg cctgctcaag gacgacccgg ccctgctccc ccgggccgtc 900
catgagctga tgcgctggtg cgggccggtg cagatgacgc agctgcgcta cgcggccgcc 960
gacgtcgacc tcgccggtac gcggatccac aagggcgacg ccgtacaact cctcctggtt 1020
gcggcgaact tcgacccccg ccactacacc gaccccgacc gtctcgatct gacgcgtcac 1080
cccgccggcc acgccgagaa ccatgtgggt ttcggccacg gtgcgcatta ctgcctgggt 1140
gccaccctcg ccaagcagga gggcgaagtc gcgttcggca agctgctcgc gcactacccg 1200
gagatgtccc tgggcatcga accggaacgt ctggagcgat tgccgctgcc tggcaactgg 1260
cggctgaatt ccctgccgtt gcggctgggg tga 1293

<210> 28
<211> 430
<212> PRT
<213> Streptomyces platensis

<400> 28
Met Ser Ala Leu Ser Ser Ser Pro Phe Ala Ala His Val Gly Lys His
1 5 10 15
Pro Gly Glu Pro Asn Val Met Glu Pro Ala Leu Leu Thr Asp Pro Phe
20 25 30
Ala Gly Tyr Gly Ala Leu Arg Glu Gln Ala Pro Val Val Arg Gly Arg
35 40 45
Phe Val Asp Asp Ser Pro Val Trp Phe Val Thr Arg Phe Glu Glu Val
50 55 60
Arg Gln Val Leu Arg Asp Gln Arg Phe Val Asn Asn Pro Ala Ala Pro
65 70 75 80
Pro Leu Ala Pro Ser Ala Glu Glu Asn Pro Leu Thr Arg Leu Met Asp
85 90 95
Met Leu Gly Leu Pro Glu His Leu Arg Val Tyr Met Leu Gly Ser Ile
100 105 110
Leu Asn Tyr Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
115 120 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu
130 135 140
Gln Ile Ala Asp Glu Leu Leu Ala Arg Leu Pro Glu Tyr Ala Glu Asp

```
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Ala Asp Arg Pro Gln Trp
            180                 185                 190
Arg Lys Trp Gly Ala Asp Leu Ile Ser Met Asp Pro Asp Arg Leu Gly
        195                 200                 205
Ala Thr Phe Pro Ala Met Ile Glu His Ile His Glu Met Val Arg Glu
    210                 215                 220
Arg Arg Ala Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Gly Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Ile Leu Thr Leu Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Ser Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val Gln Met Thr Gln Leu Arg Tyr Ala Ala Ala
305                 310                 315                 320
Asp Val Asp Leu Ala Gly Thr Arg Ile His Lys Gly Asp Ala Val Gln
                325                 330                 335
Leu Leu Leu Val Ala Ala Asn Phe Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Phe Gly Lys Leu Leu Ala His Tyr Pro
385                 390                 395                 400
Glu Met Ser Leu Gly Ile Glu Pro Glu Arg Leu Glu Arg Leu Pro Leu
                405                 410                 415
Pro Gly Asn Trp Arg Leu Asn Ser Leu Pro Leu Arg Leu Gly
            420                 425                 430
```

<210> 29
<211> 1293
<212> DNA
<213> Streptomyces lydicus

<400> 29

```
atgtcggcat tacccagcaa cacgttcacc gagcacgtcg gcaagcaccc gggcgaaccg     60
aacgtgatgg atccggcgct gatcggtgat ccgttcgccg gttacggcgc gctgcgcgag    120
cagggcccgg tcgtgcgggg gcggttcgtg gacgactccc ccgtgtggtt cgtgacccgc    180
ttcgaggagg tccgcgaggt cctgcgggac cagcggttcc ggaacaatcc ggtctcctcg    240
gcgccggacg cggaccccga ggacaccccg ctgtcccggc tgatggacat gatgggtttc    300
cccgagcacc tgcgcgtcta tctgctcggc tcgatcctca acaacgacgc ccccgaccac    360
acccggctgc gccgcctggt ctcccgggcc ttcaccgcgc ggaagatcac cgatctgcgg    420
ccgcgcgtcg cacagatagc cgacgagctg ctggcccggc tgccggagca cgccgaggac    480
ggcgtcgtcg acctgatcca gcacttcgcc tatcccctgc cgatcaccgt catctgcgaa    540
ctggtcggca tccccgagga ggaccgcccg cagtggcgca cctggggcgc cgacctggtc    600
tcgctgcagc cggaccggat gagccggtcc ttcccggcga tgatcgacca catccacgag    660
```

```
ctgatcgcgg cgcggcgccg ggcgctcacc gacgacctgc tcagcgagct gatccgaacc    720
catgacgacg acggcagccg gctcagcgac gtcgagatgg tcaccatggt cctcaccgtc    780
gtcctggccg gccacgagac caccgcgcac ctcatcggca acggcacggc ggccctgctc    840
acccaccccg accagctgcg gctgctcaag gacgacccgg cgctgctgcc gcgcgcggtg    900
cacgagttga tgcgctggtg cggcccggtg cacatgaccc agctgcgcta cgccgccgag    960
gacgtcgagc tggcgggcgt ccggatccgc aaggggggacg ccgtccagct catcctggtg   1020
tcggcgaacc gcgatccgcg ccactacacc gaacccgacc gtctggacct gacccggcac   1080
cccgccggcc acgccgagaa ccatgtgggg ttcggccacg gggcgcacta ctgtctgggc   1140
gccacgctcg ccaagcagga gggcgaggtc gccctcggcg ccctgctcag gcacttcccc   1200
gagctgtcgc tggccgtcgc gccggacgcc ctggagcgca caccggtacc gggcagctgg   1260
cggctgaatg cgctgccgct gcgtctgcgc tga                                1293
```

<210> 30
<211> 430
<212> PRT
<213> Streptomyces lydicus

<400> 30

```
Met Ser Ala Leu Pro Ser Asn Thr Phe Thr Glu His Val Gly Lys His
1               5                  10                  15
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Gly Asp Pro Phe
            20                  25                  30
Ala Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Val Asp Asp Ser Pro Val Trp Phe Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Glu Val Leu Arg Asp Gln Arg Phe Arg Asn Asn Pro Val Ser Ser
65                  70                  75                  80
Ala Pro Asp Ala Asp Pro Glu Asp Thr Pro Leu Ser Arg Leu Met Asp
                85                  90                  95
Met Met Gly Phe Pro Glu His Leu Arg Val Tyr Leu Leu Gly Ser Ile
            100                 105                 110
Leu Asn Asn Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Ala
    130                 135                 140
Gln Ile Ala Asp Glu Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Glu Asp Arg Pro Gln Trp
            180                 185                 190
Arg Thr Trp Gly Ala Asp Leu Val Ser Leu Gln Pro Asp Arg Met Ser
        195                 200                 205
Arg Ser Phe Pro Ala Met Ile Asp His Ile His Glu Leu Ile Ala Ala
    210                 215                 220
Arg Arg Arg Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Ser Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Val Leu Thr Val Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
```

```
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val His Met Thr Gln Leu Arg Tyr Ala Ala Glu
305                 310                 315                 320
Asp Val Glu Leu Ala Gly Val Arg Ile Arg Lys Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Arg Asp Pro Arg His Tyr Thr Glu Pro
            340                 345                 350
Asp Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Leu Gly Ala Leu Leu Arg His Phe Pro
385                 390                 395                 400
Glu Leu Ser Leu Ala Val Ala Pro Asp Ala Leu Glu Arg Thr Pro Val
                405                 410                 415
Pro Gly Ser Trp Arg Leu Asn Ala Leu Pro Leu Arg Leu Arg
            420                 425                 430
```

<210> 31
<211> 1293
<212> DNA
<213> Streptomyces lydicus

<400> 31

```
atgtcggcat cgaccagctc tcccctcagc gcccacgtcg gcaagcaccc gggcgaaccc     60
catgtgatgg atccggcgct gatcagcgat ccgttcggcg gctacggtgc cctgcgcgag    120
cagggaccgg tcgtccgcgg acggttcttc gacgactcgc ccttgtggtt agtgacccgc    180
ttcgaggaag tccgccaggt cctgcgcgac cagcggttcg tgaacaaccc cgccgacccg    240
gcgctcggcg tcgcgccgga ggactccccg cagctgcgcg cgctggcgat gctgggcatc    300
cccgagcacc tgcacggcta tctgctcaac tcgatcctca actacgacgc ccccgaccac    360
acccggctgc gccgcctggt ctcccgcgcc ttcaccgccc gcaagatcac cgatcttcgg    420
ccgcgggtgg cgcagataac cgccgagctg ctggaccgac tcccggagca cgccgaggac    480
ggcgtggtcg acctgatcga gcacttcgcc tacccgctgc cgatcacggt gatctgcgaa    540
cttgtcggca tcgccgcgga ggaccggccc cagtggcgtt cctggggcgc cgacctggtc    600
tcggtggacc ccgaccggct cggccggacc ttcccggcga tgatcgacca catccacgcg    660
ctgatcggcc agcggcgggc cgcgctcacc gacgacctgc tcagcgagct gatccggacc    720
catgacgacg acggcagccg gctcagcgac gtcgagatgg tcaccctggt cctcaccctc    780
gtgctggccg gccacgagac caccgcacac ctcatcggca acggcaccgc ggccctgctc    840
acccaccccg accagctgcg gctgctcaag gacgacccgg cgctgctgcc gcgcgccgtc    900
cacgagctga tgcgctggtg cgggccggtg cacgtcaccc agctgcggta cgccgccgag    960
gacgtcgacc tcgccggcac ccggatccgc aggggcgacg ccgtgcaggc cgtcctggtc   1020
tcggcgaacc acgacccgcg ccactacacc gaccccgaac gcctggacct gacccggcag   1080
cccgcgggcc gcgccgagaa ccacgtgggc ttcgggcacg gggcgcacta ctgcctgggc   1140
gccagcctcg ccaggcagga gggtgaggtc gccctgggcg ccctgttcga ccgctacccc   1200
gacctggcgc tggcggtggc gcccgaggag ctggagcgca ccccggtgcc cggtacctgg   1260
cggctgacgt cgctgccggt gcgcctgggc tga                                1293
```

<210> 32
<211> 430
<212> PRT
<213> Streptomyces lydicus

<400> 32

```
Met Ser Ala Ser Thr Ser Ser Pro Leu Ser Ala His Val Gly Lys His
1               5                  10                  15
Pro Gly Glu Pro His Val Met Asp Pro Ala Leu Ile Ser Asp Pro Phe
            20                  25                  30
Gly Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg
        35                  40                  45
Phe Phe Asp Asp Ser Pro Leu Trp Leu Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Gln Val Leu Arg Asp Gln Arg Phe Val Asn Asn Pro Ala Asp Pro
65                  70                  75                  80
Ala Leu Gly Val Ala Pro Glu Asp Ser Pro Gln Leu Arg Ala Leu Ala
                85                  90                  95
Met Leu Gly Ile Pro Glu His Leu His Gly Tyr Leu Leu Asn Ser Ile
            100                 105                 110
Leu Asn Tyr Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Ala
    130                 135                 140
Gln Ile Thr Ala Glu Leu Leu Asp Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Glu His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Ala Ala Glu Asp Arg Pro Gln Trp
            180                 185                 190
Arg Ser Trp Gly Ala Asp Leu Val Ser Val Asp Pro Asp Arg Leu Gly
        195                 200                 205
Arg Thr Phe Pro Ala Met Ile Asp His Ile His Ala Leu Ile Gly Gln
    210                 215                 220
Arg Arg Ala Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Ser Arg Leu Ser Asp Val Glu Met Val Thr Leu
                245                 250                 255
Val Leu Thr Leu Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Leu Lys Asp Asp Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met
    290                 295                 300
Arg Trp Cys Gly Pro Val His Val Thr Gln Leu Arg Tyr Ala Ala Glu
305                 310                 315                 320
Asp Val Asp Leu Ala Gly Thr Arg Ile Arg Arg Gly Asp Ala Val Gln
                325                 330                 335
Ala Val Leu Val Ser Ala Asn His Asp Pro Arg His Tyr Thr Asp Pro
            340                 345                 350
Glu Arg Leu Asp Leu Thr Arg Gln Pro Ala Gly Arg Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Ala His Tyr Cys Leu Gly Ala Ser Leu Ala
    370                 375                 380
Arg Gln Glu Gly Glu Val Ala Leu Gly Ala Leu Phe Asp Arg Tyr Pro
385                 390                 395                 400
Asp Leu Ala Leu Ala Val Ala Pro Glu Glu Leu Glu Arg Thr Pro Val
                405                 410                 415
```

```
Pro Gly Thr Trp Arg Leu Thr Ser Leu Pro Val Arg Leu Gly
            420                 425                 430


<210>  33
<211>  1281
<212>  DNA
<213>  Streptomyces tubercidicus

<400>  33
atgaactctc cgttcgccgc gcacgtcggg aaacacccgg gcgagccgaa tgtgatggac     60
cccgccctga tcaccgaccc gttcaccggc tacggcgcgc tgcgtgagca gggcccggtc    120
gtacggggcc ggttcatgga cgactcgccc gtctggctgg tgacgcggtt cgaggaggtc    180
cgccaggtcc tgcgcgacca gcggttcgtg aacaatccgg cctcgccgtc cctgaactac    240
gcgcccgagg acaacccgct gacccggctg atggagatgc tgggcctccc cgagcacctc    300
cgcgtctacc tgctcggatc gatcctcaac tacgacgccc ccgaccacac ccggctgcgc    360
cgtctggtgt cgcgggcgtt cacggcccgc aagatcaccg acctgcggcc ccgggtcgag    420
cagatcgccg acgcgctgct ggcccggctg cccgagcacg ccgaggacgg cgtcgtcgac    480
ctcatccagc acttcgccta ccccctgccg atcaccgtca tctgcgaact ggtcggcata    540
cccgaagcgg accgcccgca gtggcgaacg tggggcgccg acctcatctc gatggatccg    600
gaccggctcg gcgcctcgtt cccggcgatg atcgagcaca tccatcagat ggtccgggaa    660
cggcgcgagg cgctcaccga cgacctgctc agcgaactga tccgcaccca tgacgacgac    720
ggcgggcggc tcagcgacgt cgagatggtc accatgatcc tcacgctcgt cctcgccggc    780
cacgagacca ccgcccacct catcagcaac ggcacggcgg cgctgctcac ccaccccgac    840
cagctgcgtc tggtcaagga cgatccggcc ctcctccccc gtgccgtcca cgagctgatg    900

cgctggtgcg ggccggtgca catgacccag ctgcgctacg ccaccgccga cgtcgacctc    960
gccggcacac cgatccgcca gggcgatgcc gttcaactca tcctggtatc ggccaacttc   1020
gacccccgtc actacaccga ccccgaccgc ctcgatctca cccggcaccc cgcgggccac   1080
gccgagaacc atgtgggttt cggccatgga gcgcactact gcctgggcgc cacactcgcc   1140
aaacaggaag gtgaagtcgc cttcggcaaa ctgctcacgc actacccgga catatcgctg   1200
ggcatcgccc cggaacacct ggagcggaca ccgctgccgg gcaactggcg gctgaactcg   1260
ctgccggtgc ggttggggtg a                                             1281


<210>  34
<211>  426
<212>  PRT
<213>  Streptomyces tubercidicus

<400>  34
Met Asn Ser Pro Phe Ala Ala His Val Gly Lys His Pro Gly Glu Pro
1               5                  10                  15
Asn Val Met Asp Pro Ala Leu Ile Thr Asp Pro Phe Thr Gly Tyr Gly
            20                  25                  30
Ala Leu Arg Glu Gln Gly Pro Val Val Arg Gly Arg Phe Met Asp Asp
        35                  40                  45
Ser Pro Val Trp Leu Val Thr Arg Phe Glu Glu Val Arg Gln Val Leu
    50                  55                  60
Arg Asp Gln Arg Phe Val Asn Asn Pro Ala Ser Pro Ser Leu Asn Tyr
65                  70                  75                  80
Ala Pro Glu Asp Asn Pro Leu Thr Arg Leu Met Glu Met Leu Gly Leu
                85                  90                  95
Pro Glu His Leu Arg Val Tyr Leu Leu Gly Ser Ile Leu Asn Tyr Asp
```

```
            100                 105                 110
Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser Arg Ala Phe Thr
        115                 120                 125
Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu Gln Ile Ala Asp
    130                 135                 140
Ala Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp Gly Val Val Asp
145                 150                 155                 160
Leu Ile Gln His Phe Ala Tyr Pro Leu Pro Ile Thr Val Ile Cys Glu
                165                 170                 175
Leu Val Gly Ile Pro Glu Ala Asp Arg Pro Gln Trp Arg Thr Trp Gly
            180                 185                 190
Ala Asp Leu Ile Ser Met Asp Pro Asp Arg Leu Gly Ala Ser Phe Pro
        195                 200                 205
Ala Met Ile Glu His Ile His Gln Met Val Arg Glu Arg Arg Glu Ala
    210                 215                 220
Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr His Asp Asp Asp
225                 230                 235                 240
Gly Gly Arg Leu Ser Asp Val Glu Met Val Thr Met Ile Leu Thr Leu
                245                 250                 255
Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile Ser Asn Gly Thr
            260                 265                 270
Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu Val Lys Asp Asp
        275                 280                 285
Pro Ala Leu Leu Pro Arg Ala Val His Glu Leu Met Arg Trp Cys Gly
    290                 295                 300
Pro Val His Met Thr Gln Leu Arg Tyr Ala Thr Ala Asp Val Asp Leu
305                 310                 315                 320
Ala Gly Thr Pro Ile Arg Gln Gly Asp Ala Val Gln Leu Ile Leu Val
                325                 330                 335
Ser Ala Asn Phe Asp Pro Arg His Tyr Thr Asp Pro Asp Arg Leu Asp
            340                 345                 350
Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His Val Gly Phe Gly
        355                 360                 365
His Gly Ala His Tyr Cys Leu Gly Ala Thr Leu Ala Lys Gln Glu Gly
    370                 375                 380

Glu Val Ala Phe Gly Lys Leu Leu Thr His Tyr Pro Asp Ile Ser Leu
385                 390                 395                 400
Gly Ile Ala Pro Glu His Leu Glu Arg Thr Pro Leu Pro Gly Asn Trp
                405                 410                 415
Arg Leu Asn Ser Leu Pro Val Arg Leu Gly
            420                 425
```

<210> 35
<211> 195
<212> DNA
<213> Streptomyces tubercidicus

<400> 35
atgcggatca cgatcgacac cgacatctgt atcggcgccg gccagtgcgc cctgaccgcg 60
ccgggagtgt tcacccagga cgacgacggc ttcagcgccc tgctgcccgg ccgcgaggac 120
ggtgcgggcg acccgctggt gcgggaggcc gcccgcgcct gcccggtgca ggccatcacg 180
gtcacggacg actga 195

<210> 36
<211> 64
<212> PRT
<213> Streptomyces tubercidicus

<400> 36
Met Arg Ile Thr Ile Asp Thr Asp Ile Cys Ile Gly Ala Gly Gln Cys
1 5 10 15
Ala Leu Thr Ala Pro Gly Val Phe Thr Gln Asp Asp Asp Gly Phe Ser
20 25 30
Ala Leu Leu Pro Gly Arg Glu Asp Gly Ala Gly Asp Pro Leu Val Arg
35 40 45
Glu Ala Ala Arg Ala Cys Pro Val Gln Ala Ile Thr Val Thr Asp Asp
50 55 60

<210> 37
<211> 195
<212> DNA
<213> Streptomyces tubercidicus

<400> 37
atgcggatca ccatcgacac cgacatctgc atcggcgccg gccagtgcgc cctgaccgcg 60
ccgggagtct tcacccagga cgacgacggt ttcagcgccc tgctgcccgg ccgcgaggac 120
ggcgcgggcg acccgctggt gcgcgaggcc gcccgcgcct gccccgtgca ggccatttcg 180
gtcacggacg actga 195

<210> 38
<211> 64
<212> PRT
<213> Streptomyces tubercidicus

<400> 38
Met Arg Ile Thr Ile Asp Thr Asp Ile Cys Ile Gly Ala Gly Gln Cys
1 5 10 15
Ala Leu Thr Ala Pro Gly Val Phe Thr Gln Asp Asp Asp Gly Phe Ser
20 25 30
Ala Leu Leu Pro Gly Arg Glu Asp Gly Ala Gly Asp Pro Leu Val Arg
35 40 45
Glu Ala Ala Arg Ala Cys Pro Val Gln Ala Ile Ser Val Thr Asp Asp
50 55 60

<210> 39
<211> 9
<212> PRT
<213> Artificial Sequence
<220>
<223> Synthetic peptide.

<400> 39

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

<210> 40
<211> 10
<212> PRT
<213> Artificial Sequence
<220>
<223> Synthetic peptide.

<400> 40
Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 41
<211> 11
<212> PRT
<213> Artificial Sequence
<220>
<223> Synthetic peptide.

<400> 41
Gln Pro Glu Leu Ala Pro Glu Asp Pro Glu Asp
1 5 10

<210> 42
<211> 11
<212> PRT
<213> Artificial Sequence
<220>
<223> Synthetic peptide.

<400> 42
Tyr Thr Asp Ile Glu Met Asn Arg Leu Gly Lys
1 5 10

<210> 43
<211> 7
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 43
Ile Ala Gly His Glu Thr Thr
1 5

<210> 44
<211> 20
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (6)..(18)
<223> Nucleotides 6, 9 and 18 are "s" wherein "s" = g or c.

<400> 44
atcgcsggsc acgagacsac 20

<210> 45
<211> 7
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 45
Val Ala Gly His Glu Thr Thr
1 5

<210> 46
<211> 20
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (3)..(18)
<223> Nucleotides 3, 6, 9, and 18 are "s" wherein "s" = g or c.

<400> 46
gtsgcsggsc acgagacsac 20

<210> 47
<211> 7
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 47
Leu Ala Gly His Glu Thr Thr
1 5

<210> 48
<211> 20
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (3)..(18)
<223> Nucleotides 3, 6, 9, and 18 are "s" wherein "s" = g or c.

<400> 48
ctsgcsggsc acgagacsac 20

<210> 49
<211> 9
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 49
Leu Leu Leu Ile Ala Gly His Glu Thr
1 5

<210> 50
<211> 25
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (2)..(17)
<223> Nucleotides 2, 5, 8, 14, and 17 are "s" wherein "s" = g or c.

<400> 50
tsctsctsat cgcsggscac gagac 25

<210> 51
<211> 9
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 51
His Gln Cys Leu Gly Gln Asn Leu Ala
1 5

<210> 52
<211> 26
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (12)..(24)
<223> Nucleotides 12, 15, and 24 are "s" wherein "s" = g or c.

<400> 52
gtggtcacgg asccstgctt ggascg 26

<210> 53
<211> 8
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 53
Phe Gly His Gly Val His Gln Cys
1 5

<210> 54
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (6)..(15)
<223> Nucleotides 6, 12, and 15 are "s" wherein "s" = g or c.

<400> 54
aagccsgtgc cscasgtggt cacg 24

<210> 55
<211> 8
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 55

Phe Gly Phe Gly Val His Gln Cys
1 5

<210> 56
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (6)..(15)
<223> Nucleotides 6, 12, and 15 are "s" wherein "s" = g or c.

<400> 56
aaggcsaagc cscasgtggt cacg 24

<210> 57
<211> 8
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 57
Phe Gly His Gly Ile His Gln Cys
1 5

<210> 58
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (6)..(12)
<223> Nucleotides 6 and 12 are "s" wherein "s" = g or c.

<400> 58
aagccsgtgc cstaggtggt cacg 24

<210> 59
<211> 8
<212> PRT
<213> Streptomyces
<220>
<221> misc_feature
<222>
<223> Streptomyces consensus sequence

<400> 59
Phe Gly His Gly Val His Phe Cys
1 5

<210> 60
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (6)..(15)
<223> Nucleotides 6, 12, and 15 are "s" wherein "s" = g or c.

<400> 60
aagccsgtgc cscasgtgaa gacg 24

<210> 61
<211> 24
<212> PRT
<213> Streptomyces tubercidicus

<400> 61
His Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Thr Asp Pro
1 5 10 15
Phe Thr Gly Tyr Gly Ala Leu Arg
20

<210> 62
<211> 21
<212> PRT
<213> Streptomyces tubercidicus

<400> 62
Phe Val Asn Asn Pro Ala Ser Pro Ser Leu Asn Tyr Ala Pro Glu Asp
1 5 10 15
Asn Pro Leu Thr Arg
20

<210> 63
<211> 19
<212> PRT
<213> Streptomyces tubercidicus

<400> 63
Leu Leu Thr His Tyr Pro Asp Ile Ser Leu Gly Ile Ala Pro Glu His
1 5 10 15
Leu Glu Arg

<210> 64
<211> 17
<212> PRT
<213> Streptomyces tubercidicus

<400> 64
Val Tyr Leu Leu Gly Ser Ile Leu Asn Tyr Asp Ala Pro Asp His Thr
1 5 10 15
Arg

<210> 65
<211> 13
<212> PRT
<213> Streptomyces tubercidicus

<400> 65
Thr Trp Gly Ala Asp Leu Ile Ser Met Asp Pro Asp Arg
1 5 10

<210> 66
<211> 13
<212> PRT
<213> Streptomyces tubercidicus

<400> 66
Glu Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg
1 5 10

<210> 67
<211> 12
<212> PRT
<213> Streptomyces tubercidicus

<400> 67
Phe Met Asp Asp Ser Pro Val Trp Leu Val Thr Arg
1 5 10

<210> 68
<211> 12
<212> PRT
<213> Streptomyces tubercidicus

<400> 68
Leu Met Glu Met Leu Gly Leu Pro Glu His Leu Arg
1 5 10

<210> 69

<211> 11
<212> PRT
<213> Streptomyces tubercidicus

<400> 69
Val Glu Gln Ile Ala Asp Ala Leu Leu Ala Arg
1 5 10

<210> 70
<211> 11
<212> PRT
<213> Streptomyces tubercidicus

<400> 70
Leu Val Lys Asp Asp Pro Ala Leu Leu Pro Arg
1 5 10

<210> 71
<211> 8
<212> PRT
<213> Streptomyces tubercidicus

<400> 71
Asp Asp Pro Ala Leu Leu Pro Arg
1 5

<210> 72
<211> 8
<212> PRT
<213> Streptomyces tubercidicus

<400> 72
Thr Pro Leu Pro Gly Asn Trp Arg
1 5

<210> 73
<211> 7
<212> PRT
<213> Streptomyces tubercidicus

<400> 73
Leu Asn Ser Leu Pro Val Arg
1 5

<210> 74
<211> 7
<212> PRT
<213> Streptomyces tubercidicus

<400> 74
Ile Thr Asp Leu Arg Pro Arg
1 5

<210> 75
<211> 7
<212> PRT
<213> Streptomyces tubercidicus

<400> 75
Glu Gln Gly Pro Val Val Arg
1 5

<210> 76
<211> 7
<212> PRT
<213> Streptomyces tubercidicus

<400> 76
Ala Val His Glu Leu Met Arg
1 5

<210> 77
<211> 5
<212> PRT
<213> Streptomyces tubercidicus

<400> 77
Ala Phe Thr Ala Arg
1 5

<210> 78
<211> 5
<212> PRT
<213> Streptomyces tubercidicus

<400> 78
Phe Glu Glu Val Arg
1 5

<210> 79
<211> 7
<212> PRT
<213> Streptomyces tubercidicus

<400> 79
Pro Gly Glu Asp Asn Val Met
1 5

<210> 80
<211> 21
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (3)..(18)
<223> Nucleotides 3, 6, 12, and 18 are "s" wherein "s" = c or g.
<220>
<221> misc_feature
<222> (9)..(9)
<223> Nucleotide 9 is "r" wherein "r" = a or g.
<220>
<221> misc_feature
<222> (15)..(15)
<223> Nucleotide 15 is "y" wherein "y" = c or t.

<400> 80
ccsggsgarc csaaygtsat g 21

<210> 81
<211> 7
<212> PRT
<213> Streptomyces tubercidicus

<400> 81
Ala Leu Ile Thr Asp Pro Phe
1 5
<210> 82
<211> 21
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (3)..(18)
<223> Nucleotides 3, 6, 12, and 18 are "s"wherein "s" = c or g.

<400> 82
gcsctsatya csgacccstt c 21

<210> 83
<211> 8
<212> PRT
<213> Streptomyces tubercidicus

<400> 83
Phe Met Asp Asp Ser Pro Val Trp
1 5

<210> 84
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (13)..(13)
<223> Nucleotide 13 is "w" wherein "w" = a or t.
<220>
<221> misc_feature
<222> (14)..(21)
<223> Nucleotides 14, 15, 18, and 21 are "s" wherein "s" = c or g.

<400> 84
ttcatggacg acwssccsgt stgg 24

<210> 85
<211> 8
<212> PRT
<213> Streptomyces tubercidicus

<400> 85
Leu Asn Tyr Asp Ala Pro Asp His
1 5

<210> 86
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (3)..(18)
<223> Nucleotides 3, 15 and 18 are "s" wherein "s" = c or g.
<220>
<221> misc_feature
<222> (6)..(9)
<223> Nucleotides 6 and 9 are "y" wherein "y" = c or t.

<400> 86
ctsaaytayg acgcsccsga ccac 24

<210> 87
<211> 8
<212> PRT
<213> Streptomyces tubercidicus

<400> 87
Val Glu Gln Ile Ala Asp Ala Leu
1 5

<210> 88
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (3)..(24)
<223> Nucleotides 3, 15, 21, and 24 are "s" wherein "s" = c or g.
<220>
<221> misc_feature
<222> (12)..(12)
<223> Nucleotide 12 is "y" wherein "y" = c or t.
<220>
<221> misc_feature
<222> (6)..(6)
<223> Nucleotide 6 is "r" wherein "r" = a or g.

<400> 88
gtsgarcaga tygcsgacgc scts 24

<210> 89
<211> 8
<212> PRT
<213> Streptomyces tubercidicus

<400> 89
Asp Leu Ile Ser Met Asp Pro Asp
1 5

<210> 90
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<221> misc_feature
<222> (6)..(21)
<223> Nucleotides 6, 11, 12, and 21 are "s" wherein "s" = c or g.
<220>
<221> misc_feature
<222> (9)..(9)
<223> Nucleotide 9 is "r" wherein "r" = a or g.
<220>
<221> misc_feature
<222> (10)..(10)
<223> Nucleotide 10 is "w" wherein "w" = a or t.

<400> 90
ctggastarw sstacctggg sctg 24

<210> 91
<211> 36

<212> DNA
<213> Streptomyces tubercidicus

<400> 91
agattaatta atgtcggaat taatgaactg tccgtt 36

<210> 92
<211> 32
<212> DNA
<213> Streptomyces tubercidicus

<400> 92
aaactcaccc caaccgcacc ggcagcgagt tc 32

<210> 93
<211> 7
<212> PRT
<213> Streptomyces tubercidicus

<400> 93
Met Ser Glu Leu Met Asn Ser
1 5

<210> 94
<211> 1293
<212> DNA
<213> Streptomyces tubercidicus

<400> 94
atgtcggcaa tatccagctc cccgttcgcc gcacacgtcg gaaagcatcc cggcgagccg 60
aatgtgatgg acccggcgct gatcaccgac ccgttcggcg gctacggcgc actgcgtgag 120
caaggccccg tcctaccggg ccggttcatg gacgactcac ccgtctggct cgtgacgcgc 180
ttcgaagagg tccgccaagt cctgcgcgat cagcggttcc tgaacaaccc ggccgcgtcg 240
tcaccggggc attcgatcga cgagagcccc acggccaggc tgctggacat gatggggatg 300
cccgaacatt tccggccgta tctgatgggg tcgatcctca acaacgacgc ccccgaccac 360
acccggctgc gccgtctggt gtcacgcgcg ttcacggcac gcaagatcac cgatctgcgg 420
ccgcgggtcg agcagctcgc cgacgagctg ctggcccggc ttcccgagca cgccgaggac 480
ggtgtggtcg acctgatcaa gcacttcgcc tatcccctgc cgatcaccgt gatctgcgaa 540
ctggtcggca tcccggaagc ggaccgcccg caatggcgga agtggggcgc cgacctcgtt 600
tcgctgcagc cggagcggct cagcacctcg ttcccggcga tgatcgagca catccatgaa 660
ctgatccgcg agcggcgcgg cgcgctcacc gacgatctgc tcagcgagct gatccgtacc 720
catgacgacg acggcagccg gctcagcgac gtcgagatgg tcaccatggt cctcaccgtc 780
gtcctggccg gccacgagac caccgcccac ctgataggca acggcacggc ggcgctgctc 840
acccaccccg accagctgcg cctggtcaag gacgacccgg agctgcttcc gcgtgccgtc 900
cacgagctgc tgcgctggtg cgggccggtc cagatgaccc agctgcggta cgcctccgag 960
gatgtcgaga tcgccgggac gccgatccgt aagggcgacg ccgtacaact catcctggta 1020
tcggcgaact tcgaccccg ccactacacc gcccccgaac gcctcgacct gacccgccac 1080
cccgccggcc acgccgagaa ccatgtgggc ttcggccacg gaatgcacta ctgcctgggc 1140
gccaccctcg ccaaacagga gggcgaagtc gcgttcggca agctcttcac gcactacccg 1200
gagctgtcgc tggccgtcgc accggacgag ttggagcgaa cgccggtgcc cggcagctgg 1260

cggttggatt cgctgccggt gcggttgggg tga 1293

<210> 95
<211> 430
<212> PRT
<213> Streptomyces tubercidicus

<400> 95

```
Met Ser Ala Ile Ser Ser Ser Pro Phe Ala Ala His Val Gly Lys His
 1               5                  10                  15
Pro Gly Glu Pro Asn Val Met Asp Pro Ala Leu Ile Thr Asp Pro Phe
            20                  25                  30
Gly Gly Tyr Gly Ala Leu Arg Glu Gln Gly Pro Val Leu Pro Gly Arg
        35                  40                  45
Phe Met Asp Asp Ser Pro Val Trp Leu Val Thr Arg Phe Glu Glu Val
    50                  55                  60
Arg Gln Val Leu Arg Asp Gln Arg Phe Leu Asn Asn Pro Ala Ala Ser
65                  70                  75                  80
Ser Pro Gly His Ser Ile Asp Glu Ser Pro Thr Ala Arg Leu Leu Asp
                85                  90                  95
Met Met Gly Met Pro Glu His Phe Arg Pro Tyr Leu Met Gly Ser Ile
            100                 105                 110
Leu Asn Asn Asp Ala Pro Asp His Thr Arg Leu Arg Arg Leu Val Ser
        115                 120                 125
Arg Ala Phe Thr Ala Arg Lys Ile Thr Asp Leu Arg Pro Arg Val Glu
    130                 135                 140
Gln Leu Ala Asp Glu Leu Leu Ala Arg Leu Pro Glu His Ala Glu Asp
145                 150                 155                 160
Gly Val Val Asp Leu Ile Lys His Phe Ala Tyr Pro Leu Pro Ile Thr
                165                 170                 175
Val Ile Cys Glu Leu Val Gly Ile Pro Glu Ala Asp Arg Pro Gln Trp
            180                 185                 190
Arg Lys Trp Gly Ala Asp Leu Val Ser Leu Gln Pro Glu Arg Leu Ser
        195                 200                 205
Thr Ser Phe Pro Ala Met Ile Glu His Ile His Glu Leu Ile Arg Glu
    210                 215                 220
Arg Arg Gly Ala Leu Thr Asp Asp Leu Leu Ser Glu Leu Ile Arg Thr
225                 230                 235                 240
His Asp Asp Asp Gly Ser Arg Leu Ser Asp Val Glu Met Val Thr Met
                245                 250                 255
Val Leu Thr Val Val Leu Ala Gly His Glu Thr Thr Ala His Leu Ile
            260                 265                 270
Gly Asn Gly Thr Ala Ala Leu Leu Thr His Pro Asp Gln Leu Arg Leu
        275                 280                 285
Val Lys Asp Asp Pro Glu Leu Leu Pro Arg Ala Val His Glu Leu Leu
    290                 295                 300
Arg Trp Cys Gly Pro Val Gln Met Thr Gln Leu Arg Tyr Ala Ser Glu
305                 310                 315                 320
Asp Val Glu Ile Ala Gly Thr Pro Ile Arg Lys Gly Asp Ala Val Gln
                325                 330                 335
Leu Ile Leu Val Ser Ala Asn Phe Asp Pro Arg His Tyr Thr Ala Pro
            340                 345                 350
```

```
Glu Arg Leu Asp Leu Thr Arg His Pro Ala Gly His Ala Glu Asn His
        355                 360                 365
Val Gly Phe Gly His Gly Met His Tyr Cys Leu Gly Ala Thr Leu Ala
    370                 375                 380
Lys Gln Glu Gly Glu Val Ala Phe Gly Lys Leu Phe Thr His Tyr Pro
385                 390                 395                 400
Glu Leu Ser Leu Ala Val Ala Pro Asp Glu Leu Glu Arg Thr Pro Val
                405                 410                 415
Pro Gly Ser Trp Arg Leu Asp Ser Leu Pro Val Arg Leu Gly
            420                 425                 430
```

<210> 96
<211> 18
<212> DNA
<213> Artificial Sequence

<400> 96
cgsccscccsc tswssaas 18

<210> 97
<211> 21
<212> DNA
<213> Artificial Sequence

<400> 97
sassgcstts bcccartgyt c 21

<210> 98
<211> 1266
<212> DNA
<213> Streptomyces tubercidicus

<400> 98
gtggtcgacg cacaccagac gttcgtcatc gtcgggggtg gcctggccgg cgcaaaggcc 60
gcggagactc tccgcgcgga ggggttcacc ggccgggtga tcctcatctg tgacgagcgc 120
gaccacccgt acgagcgccc cccgctctcc aaggggttcc tgctcggcaa ggaagagcgc 180
gacagcgtgt tcgtccatga gcccgcctgg tacgcccagg cacagatcga actgcacctg 240
ggccagcccg ccgtccgcct cgaccccgag ggcaggaccg tccgcctcgg cgacggcacc 300
ctgatcgcct acgacaagct gctgctggcc accggcgccg aaccgcggcg cctggacatc 360
cccggcaccg gcctggccgg cgtgcaccac ctgcgccgcc tcgcccacgc cgaacggctg 420
cgcggcgtcc tggcctccct cggccgcgac aacggccatc tggtgatcgc cggagccggc 480
tggatcggcc tggaggtcgc cgccgcggcc cgctcctacg gcgccgaggt gaccgtcgtc 540
gaggccgccc cgacgccgct gcacggcatc ctggggcccg aactcggcgg tctgttcacc 600
gatctgcacc gcgagcacgg cgtccgcttc cacttcggcg cccgcttcac cgagatcgtc 660
ggagagggcg gcatggtgct cgccgtgcgc accgacgacg gcgaggaaca ccccgcccac 720
gatgtgctcg ccgcgatcgg cgccgccccg cgcaccgcgc tcgccgaaca ggccgggctg 780
gatctcgccg acccggagac cggcggcggg gtggccgtcg acgcggcgct gcgcacctcc 840
gacccgtaca tctacgccgc cggtgacgtc gccgccgccg accacccgct gctggacacc 900

```
cggctgcggg tcgaacactg ggccaacgcc ctcaacggcg gcccggccgc cgcccgcgcc 960
atgctcggcc aggacatcag ctacgaccgc atcccgtact tcttctccga ccagtacgac 1020
gtcggcatgg agtactccgg ctacgccccg cccggctcgt acgcccaggt cgtctgccgc 1080
ggcgacgtcg ccaagcggga gttcatcgcc ttctggctgg cggcggacgg ccggctgctc 1140
gcgggcatga acgtcaacgt ctgggacgtc gccgagtcca tccagcaact catccgctcc 1200
ggggcgccgt tggagcccgg cgcactggcc gatccgcagg ttccgctggc ggcactgctc 1260
ccgtag                                                            1266
```

<210> 99
<211> 421
<212> PRT
<213> Streptomyces tubercidicus

<400> 99

```
Val Val Asp Ala His Gln Thr Phe Val Ile Val Gly Gly Gly Leu Ala
 1               5                  10                  15
Gly Ala Lys Ala Ala Glu Thr Leu Arg Ala Glu Gly Phe Thr Gly Arg
            20                  25                  30
Val Ile Leu Ile Cys Asp Glu Arg Asp His Pro Tyr Glu Arg Pro Pro
        35                  40                  45
Leu Ser Lys Gly Phe Leu Leu Gly Lys Glu Glu Arg Asp Ser Val Phe
    50                  55                  60
Val His Glu Pro Ala Trp Tyr Ala Gln Ala Gln Ile Glu Leu His Leu
65                  70                  75                  80
Gly Gln Pro Ala Val Arg Leu Asp Pro Glu Gly Arg Thr Val Arg Leu
                85                  90                  95
Gly Asp Gly Thr Leu Ile Ala Tyr Asp Lys Leu Leu Leu Ala Thr Gly
            100                 105                 110
Ala Glu Pro Arg Arg Leu Asp Ile Pro Gly Thr Gly Leu Ala Gly Val
        115                 120                 125
His His Leu Arg Arg Leu Ala His Ala Glu Arg Leu Arg Gly Val Leu
    130                 135                 140
Ala Ser Leu Gly Arg Asp Asn Gly His Leu Val Ile Ala Gly Ala Gly
145                 150                 155                 160
Trp Ile Gly Leu Glu Val Ala Ala Ala Ala Arg Ser Tyr Gly Ala Glu
                165                 170                 175
Val Thr Val Val Glu Ala Ala Pro Thr Pro Leu His Gly Ile Leu Gly
            180                 185                 190
Pro Glu Leu Gly Gly Leu Phe Thr Asp Leu His Arg Glu His Gly Val
        195                 200                 205
Arg Phe His Phe Gly Ala Arg Phe Thr Glu Ile Val Gly Glu Gly Gly
    210                 215                 220
Met Val Leu Ala Val Arg Thr Asp Asp Gly Glu Glu His Pro Ala His
225                 230                 235                 240
Asp Val Leu Ala Ala Ile Gly Ala Ala Pro Arg Thr Ala Leu Ala Glu
                245                 250                 255
Gln Ala Gly Leu Asp Leu Ala Asp Pro Glu Thr Gly Gly Gly Val Ala
            260                 265                 270
Val Asp Ala Ala Leu Arg Thr Ser Asp Pro Tyr Ile Tyr Ala Ala Gly
        275                 280                 285
Asp Val Ala Ala Ala Asp His Pro Leu Leu Asp Thr Arg Leu Arg Val
    290                 295                 300
```

```
Glu His Trp Ala Asn Ala Leu Asn Gly Gly Pro Ala Ala Ala Arg Ala
305                 310                 315                 320
Met Leu Gly Gln Asp Ile Ser Tyr Asp Arg Ile Pro Tyr Phe Phe Ser
                325                 330                 335
Asp Gln Tyr Asp Val Gly Met Glu Tyr Ser Gly Tyr Ala Pro Pro Gly
            340                 345                 350
Ser Tyr Ala Gln Val Val Cys Arg Gly Asp Val Ala Lys Arg Glu Phe
        355                 360                 365
Ile Ala Phe Trp Leu Ala Ala Asp Gly Arg Leu Leu Ala Gly Met Asn
    370                 375                 380
Val Asn Val Trp Asp Val Ala Glu Ser Ile Gln Gln Leu Ile Arg Ser
385                 390                 395                 400
Gly Ala Pro Leu Glu Pro Gly Ala Leu Ala Asp Pro Gln Val Pro Leu
                405                 410                 415
Ala Ala Leu Leu Pro
            420
```

<210> 100
<211> 1314
<212> DNA
<213> Streptomyces tubercidicus

<400> 100

```
atgcccgctg cacgccgccg ccttcgacct ccgcaccgga gcggcgacct gcctgcccgc 60
ccgccgggcc gtgcgcaccc accccgtgac cgtccaggac ggcatgatct acgtccatca 120
cgccgcggag gagggcaccg ccgcatgaag tcggtcgctg tcatcggggc ctcgctggcg 180
ggcctgtacg ccgcgcggtc cctgcgttcc caggggttcg acggccgcct ggtgatcgtc 240
ggggacgagt gccacggccc ctacgaccgg cccccgctgt ccaaggactt cctcaccggc 300
gccaccgacc cgggccgact cgccctggcc gacgccgagg agatcgccga actcgacgcc 360
gaatggctgc tgggcacccg ggccaccggg ctcgacaccg gcggacgcac ggtgctgctc 420
gatggcggcc ggtccctgac caccgacggc gtggtcctcg ccaccggcgc cgccccgcgc 480
ctgctccccg gaccggtgcc cgccggggtc cacaccctgc gcaccctcga cgacgcccag 540
gcgctccgtg cggatctggc gccggcgccg gtccgggtcg tggtgatcgg cggcggcttc 600
atcggcgccg aggtcgcctc gtcctgcgcc gccctaggcc atgacgtcac cgtggtcgag 660
gccgcgccgc tccccctcgt cccccaactc ggccacgcca tggccgagat ctgcgccgcc 720
ctgcatgcgg accacggcgt cacgctgctc accggaaccg gtgtcgcccg gctgcgcagc 780
gagggcgacg gccggcgcgt caccggcgtc gagctgaccg acggccgcct gctccccgcc 840
gacgtggtcg tcgtcggcat cggggtacgc ccccgcaccg cctggctcac ggactccgga 900
ctgccgctcg acgacggtgt gctctgcgac gcgggctgtg tcaccccgct gcccgccgtc 960
gtggccgtcg gcgacgtcgc cagggtggac ggcgcccgtg ccgagcactg gaccagcgcc 1020
accgaacagg ccgccgtggc ggcgcggaac ctgctggccg gcagcaccgt cgcgacccac 1080
cggagcctgc cgtacttctg gtccgaccag tacggcgtcc gcatccagtt cgcgggccac 1140
cggctgccca ccgacacacc gcgcgtcctc gaaggctccc ccgacgaccg cagcttcctc 1200
gcctgttacg aacgggacgg acgcaccacc gcggtgctcg ccctcaaccg gccccgcccc 1260
ttcatgcggc tccgccgcga actcgcccgc accgccctgt cggccaccac ctga       1314
```

<210> 101
<211> 437
<212> PRT
<213> Streptomyces tubercidicus

<400> 101

```
Met Pro Ala Ala Arg Arg Arg Leu Arg Pro Pro His Arg Ser Gly Asp
 1               5                  10                  15
Leu Pro Ala Arg Pro Pro Gly Arg Ala His Pro Pro Arg Asp Arg Pro
            20                  25                  30
Gly Arg His Asp Leu Arg Pro Ser Arg Arg Gly Gly Gly His Arg Arg
        35                  40                  45
Met Lys Ser Val Ala Val Ile Gly Ala Ser Leu Ala Gly Leu Tyr Ala
    50                  55                  60
Ala Arg Ser Leu Arg Ser Gln Gly Phe Asp Gly Arg Leu Val Ile Val
65                  70                  75                  80
Gly Asp Glu Cys His Gly Pro Tyr Asp Arg Pro Pro Leu Ser Lys Asp
                85                  90                  95
Phe Leu Thr Gly Ala Thr Asp Pro Gly Arg Leu Ala Leu Ala Asp Ala
            100                 105                 110
Glu Glu Ile Ala Glu Leu Asp Ala Glu Trp Leu Leu Gly Thr Arg Ala
        115                 120                 125
Thr Gly Leu Asp Thr Gly Gly Arg Thr Val Leu Leu Asp Gly Gly Arg
    130                 135                 140
Ser Leu Thr Thr Asp Gly Val Val Leu Ala Thr Gly Ala Ala Pro Arg
145                 150                 155                 160
Leu Leu Pro Gly Pro Val Pro Ala Gly Val His Thr Leu Arg Thr Leu
                165                 170                 175
Asp Asp Ala Gln Ala Leu Arg Ala Asp Leu Ala Pro Ala Pro Val Arg
            180                 185                 190
Val Val Val Ile Gly Gly Gly Phe Ile Gly Ala Glu Val Ala Ser Ser
        195                 200                 205
Cys Ala Ala Leu Gly His Asp Val Thr Val Val Glu Ala Ala Pro Leu
    210                 215                 220
Pro Leu Val Pro Gln Leu Gly His Ala Met Ala Glu Ile Cys Ala Ala
225                 230                 235                 240
Leu His Ala Asp His Gly Val Thr Leu Leu Thr Gly Thr Gly Val Ala
                245                 250                 255
Arg Leu Arg Ser Glu Gly Asp Gly Arg Arg Val Thr Gly Val Glu Leu
            260                 265                 270
Thr Asp Gly Arg Leu Leu Pro Ala Asp Val Val Val Val Gly Ile Gly
        275                 280                 285
Val Arg Pro Arg Thr Ala Trp Leu Thr Asp Ser Gly Leu Pro Leu Asp
    290                 295                 300
Asp Gly Val Leu Cys Asp Ala Gly Cys Val Thr Pro Leu Pro Ala Val
305                 310                 315                 320
Val Ala Val Gly Asp Val Ala Arg Val Asp Gly Ala Arg Ala Glu His
                325                 330                 335
Trp Thr Ser Ala Thr Glu Gln Ala Ala Val Ala Ala Arg Asn Leu Leu
            340                 345                 350
Ala Gly Ser Thr Val Ala Thr His Arg Ser Leu Pro Tyr Phe Trp Ser
        355                 360                 365
Asp Gln Tyr Gly Val Arg Ile Gln Phe Ala Gly His Arg Leu Pro Thr
    370                 375                 380
Asp Thr Pro Arg Val Leu Glu Gly Ser Pro Asp Asp Arg Ser Phe Leu
385                 390                 395                 400
Ala Cys Tyr Glu Arg Asp Gly Arg Thr Thr Ala Val Leu Ala Leu Asn
                405                 410                 415
```

```
Arg Pro Arg Pro Phe Met Arg Leu Arg Arg Glu Leu Ala Arg Thr Ala
            420                 425                 430
Leu Ser Ala Thr Thr
        435
```

<210> 102
<211> 1233
<212> DNA
<213> Streptomyces tubercidicus

<400> 102
atggcccaga acacggcatt catcatcgcg ggagcggggc tggccggggc gaaggccgcg 60
gagacactgc gcgcggaggg cttcggcggc cccgtcctgc tgctgggcga cgagcgcgag 120
cgtccctacg agcggccgcc gctgtccaag ggctacctct tgggcacctc cgagcgggag 180
aaggcgtacg tccatccgcc ccagtggtac gccgagcacg acgtcgatct gcggctgggc 240
aacgccgtca ccgccctcga cccggccggc cacgaggtga ccctcgccga cggcagccgg 300
ctgggctacg ccaagctgct gctggccacc ggctccactc cgcgccggct gccggtgccc 360
ggcgccgacc tcgacggggt ccacacgctg cggtacctgg cggacagcga ccgcctcaag 420
gacctcttcc ggtccgcgtc ccggatcgtg gtgatcggcg gcggctggat cggcctggag 480
accacggccg ccgcgcgtgc ggcgggggtc gaggtgaccg tgctggagtc ggcgccgctg 540
cccctgctgg gggtgctggg ccgcgaggtc gcccaggtct tcgccgatct gcacaccgag 600
cacggtgtcg cgctgcgctg cgacacccag gtcacggaga tcaccggcac gaacggcgcg 660
gtcgacgggg tacggctggc cgacggcacc cggatcgcgg ccgacgcggt gatcgtcggc 720
gtcgggatca cccccaactc cgagacggcc gccgcggccg ggctcaaggt cgacaacggc 780
gtcgtcgtgg acgagcggct gtgctcctcc cacccggaca tctacgccgc cggcgacgtc 840
gccaacgcct accacccccct cctgggcaag cacctccgcg tcgagcactg ggccaacgcc 900
ctccaccagc cgaagaccgc ggcccgggcc atgctgggcg gggaggccgg ctacgaccgg 960
ctgccgtact tcttcaccga ccagtacgac ctgggcatgg agtacacggg gcatgtggag 1020
ccgggcgggt acgaccgcgt ggtgttccgc ggcgacaccg gtgcccgcga gttcatcgcc 1080
ttctggctct ccggcggccg ggtgctggcc gggatgaatg tgaacgtatg ggacgtcacc 1140
gacccgatcc gggccctggt ggcgagcggg cgggccgtgg accccgagcg gctcgccgac 1200
gcggacgtac cgctggcgga tctggtcccc tga 1233

<210> 103
<211> 410
<212> PRT
<213> Streptomyces tubercidicus

<400> 103
```
Met Ala Gln Asn Thr Ala Phe Ile Ile Ala Gly Ala Gly Leu Ala Gly
 1               5                  10                  15
Ala Lys Ala Ala Glu Thr Leu Arg Ala Glu Gly Phe Gly Gly Pro Val
            20                  25                  30
Leu Leu Leu Gly Asp Glu Arg Glu Arg Pro Tyr Glu Arg Pro Pro Leu
        35                  40                  45
Ser Lys Gly Tyr Leu Leu Gly Thr Ser Glu Arg Glu Lys Ala Tyr Val
    50                  55                  60
His Pro Pro Gln Trp Tyr Ala Glu His Asp Val Asp Leu Arg Leu Gly
65                  70                  75                  80
Asn Ala Val Thr Ala Leu Asp Pro Ala Gly His Glu Val Thr Leu Ala
```

```
                85                  90                  95
Asp Gly Ser Arg Leu Gly Tyr Ala Lys Leu Leu Leu Ala Thr Gly Ser
            100                 105                 110
Thr Pro Arg Arg Leu Pro Val Pro Gly Ala Asp Leu Asp Gly Val His
        115                 120                 125
Thr Leu Arg Tyr Leu Ala Asp Ser Asp Arg Leu Lys Asp Leu Phe Arg
    130                 135                 140
Ser Ala Ser Arg Ile Val Val Ile Gly Gly Gly Trp Ile Gly Leu Glu
145                 150                 155                 160
Thr Thr Ala Ala Ala Arg Ala Ala Gly Val Glu Val Thr Val Leu Glu
                165                 170                 175
Ser Ala Pro Leu Pro Leu Leu Gly Val Leu Gly Arg Glu Val Ala Gln
            180                 185                 190
Val Phe Ala Asp Leu His Thr Glu His Gly Val Ala Leu Arg Cys Asp
        195                 200                 205
Thr Gln Val Thr Glu Ile Thr Gly Thr Asn Gly Ala Val Asp Gly Val
    210                 215                 220
Arg Leu Ala Asp Gly Thr Arg Ile Ala Ala Asp Ala Val Ile Val Gly
225                 230                 235                 240
Val Gly Ile Thr Pro Asn Ser Glu Thr Ala Ala Ala Ala Gly Leu Lys
                245                 250                 255
Val Asp Asn Gly Val Val Val Asp Glu Arg Leu Cys Ser Ser His Pro
            260                 265                 270
Asp Ile Tyr Ala Ala Gly Asp Val Ala Asn Ala Tyr His Pro Leu Leu
        275                 280                 285
Gly Lys His Leu Arg Val Glu His Trp Ala Asn Ala Leu His Gln Pro
    290                 295                 300
Lys Thr Ala Ala Arg Ala Met Leu Gly Gly Glu Ala Gly Tyr Asp Arg
305                 310                 315                 320
Leu Pro Tyr Phe Phe Thr Asp Gln Tyr Asp Leu Gly Met Glu Tyr Thr
                325                 330                 335
Gly His Val Glu Pro Gly Gly Tyr Asp Arg Val Val Phe Arg Gly Asp
            340                 345                 350
Thr Gly Ala Arg Glu Phe Ile Ala Phe Trp Leu Ser Gly Gly Arg Val
        355                 360                 365
Leu Ala Gly Met Asn Val Asn Val Trp Asp Val Thr Asp Pro Ile Arg
    370                 375                 380
Ala Leu Val Ala Ser Gly Arg Ala Val Asp Pro Glu Arg Leu Ala Asp
385                 390                 395                 400
Ala Asp Val Pro Leu Ala Asp Leu Val Pro
                405                 410
```

<210> 104
<211> 1266
<212> DNA
<213> Streptomyces tubercidicus

<400> 104
gtggtcgacg cacaccagac gttcgtcatc gtcgggggtg gcctggccgg cgcaaaggcc 60
gcggagactc tccgcgcgga agggttcacc ggccgggtga tcctcatctg tgacgagcgc 120
gaccacccgt acgagcgccc cccgctctcc aaggggttcc tgctcggcaa ggaagagcgc 180
gacagcgttt tcgtccacga acccgcctgg tacgcccagg cacagatcga actgcacctg 240

```
ggccagcccg ccgtccgcct cgaccccgag gcgaagaccg tccgcctcgg cgacggcacc 300
ctgatcgcct acgacaagct gctgctggcc accggcgccg agccgcgccg cctggacatc 360
cccggcaccg gcctggccgg cgtgcaccac ctgcgccgcc tcgcccacgc cgaacggctg 420
cgcggcgtcc tggcctccct cgggcgggac aacgggcatc tggtgatcgc cggcgccggc 480
tggatcggcc tggaggtcgc cgccgcggcc cgctcctacg gcgccgaggt caccgtcgtc 540
gaggccgccc cgacaccgct gcacggcatc ctggggcccg aactcggcgg cctgttcacc 600
gaactgcacc gcgcacacgg cgtgcgcttc cacttcggcg cccgtttcac cgagatcgtc 660
ggacaggacg gcatggtgct cgccgtgcgc accgacgacg gcgaggagca ccccgcccac 720
gacgtgctcg ccgcgatcgg cgccgccccg cgcaccgcac tcgccgaaca ggccggactc 780
gacctcgccg acccggaggc cggcggcggc gtggccgtcg acgcgacgct gcgcacctcc 840
gacccgtaca tctacgccgc cggcgacgtg gccgccgccg accacccccct cctggacacc 900
cggctgcgcg tcgaacactg ggccaacgcc ctcaacggcg gcccggccgc cgcgcgcgcc 960
atgctcggcc aggacatcag ctacgaccgc gtcccgtact tcttctccga ccagtacgac 1020
gtcggcatgg agtactccgg ctacgccccg cccggctcct acgcacaggt cgtctgccgc 1080
ggcgacgtcg ccaaacggga gttcatcgcg ttctggctcg gcgaggacgg acggctgctc 1140
gcggggatga acgtcaacgt ctgggacgtc gccgaaacca tccagcaact catccgcggc 1200
ggggtgcggt tggagcccgg cgagctggct gatccggagg ttccgctgac ctcactgctc 1260
ccgtag                                                            1266
```

<210> 105
<211> 421
<212> PRT
<213> Streptomyces tubercidicus

<400> 105

```
Val Val Asp Ala His Gln Thr Phe Val Ile Val Gly Gly Gly Leu Ala
 1               5                  10                  15
Gly Ala Lys Ala Ala Glu Thr Leu Arg Ala Glu Gly Phe Thr Gly Arg
            20                  25                  30
Val Ile Leu Ile Cys Asp Glu Arg Asp His Pro Tyr Glu Arg Pro Pro
        35                  40                  45
Leu Ser Lys Gly Phe Leu Leu Gly Lys Glu Glu Arg Asp Ser Val Phe
    50                  55                  60
Val His Glu Pro Ala Trp Tyr Ala Gln Ala Gln Ile Glu Leu His Leu
65                  70                  75                  80
Gly Gln Pro Ala Val Arg Leu Asp Pro Glu Ala Lys Thr Val Arg Leu
                85                  90                  95
Gly Asp Gly Thr Leu Ile Ala Tyr Asp Lys Leu Leu Leu Ala Thr Gly
            100                 105                 110
Ala Glu Pro Arg Arg Leu Asp Ile Pro Gly Thr Gly Leu Ala Gly Val
        115                 120                 125
His His Leu Arg Arg Leu Ala His Ala Glu Arg Leu Arg Gly Val Leu
    130                 135                 140
Ala Ser Leu Gly Arg Asp Asn Gly His Leu Val Ile Ala Gly Ala Gly
145                 150                 155                 160
Trp Ile Gly Leu Glu Val Ala Ala Ala Ala Arg Ser Tyr Gly Ala Glu
                165                 170                 175
Val Thr Val Val Glu Ala Ala Pro Thr Pro Leu His Gly Ile Leu Gly
            180                 185                 190
Pro Glu Leu Gly Gly Leu Phe Thr Glu Leu His Arg Ala His Gly Val
        195                 200                 205
Arg Phe His Phe Gly Ala Arg Phe Thr Glu Ile Val Gly Gln Asp Gly
```

```
    210                 215                 220
Met Val Leu Ala Val Arg Thr Asp Asp Gly Glu Glu His Pro Ala His
225                 230                 235                 240
Asp Val Leu Ala Ala Ile Gly Ala Ala Pro Arg Thr Ala Leu Ala Glu
                245                 250                 255
Gln Ala Gly Leu Asp Leu Ala Asp Pro Glu Ala Gly Gly Gly Val Ala
            260                 265                 270
Val Asp Ala Thr Leu Arg Thr Ser Asp Pro Tyr Ile Tyr Ala Ala Gly
        275                 280                 285
Asp Val Ala Ala Ala Asp His Pro Leu Leu Asp Thr Arg Leu Arg Val
    290                 295                 300
Glu His Trp Ala Asn Ala Leu Asn Gly Gly Pro Ala Ala Ala Arg Ala
305                 310                 315                 320
Met Leu Gly Gln Asp Ile Ser Tyr Asp Arg Val Pro Tyr Phe Phe Ser
                325                 330                 335
Asp Gln Tyr Asp Val Gly Met Glu Tyr Ser Gly Tyr Ala Pro Pro Gly
            340                 345                 350
Ser Tyr Ala Gln Val Val Cys Arg Gly Asp Val Ala Lys Arg Glu Phe
        355                 360                 365
Ile Ala Phe Trp Leu Gly Glu Asp Gly Arg Leu Leu Ala Gly Met Asn
    370                 375                 380
Val Asn Val Trp Asp Val Ala Glu Thr Ile Gln Gln Leu Ile Arg Gly
385                 390                 395                 400
Gly Val Arg Leu Glu Pro Gly Glu Leu Ala Asp Pro Glu Val Pro Leu
                405                 410                 415
Thr Ser Leu Leu Pro
            420
```

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO
Syngenta
3054 Cornwallis Road
Research Triangle Park, NC 27709

VIABILITY STATEMENT

issued pursuant to Rule 10.2 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF THE PARTY TO WHOM THE VIABILITY STATEMENT IS ISSUED

I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM
Name: Syngenta Address: 3054 Cornwallis Road Research Triangle Park, NC 27709	Depositor's taxonomic designation and accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: *Streptomyces lividans* NRRL 30478 Date of: May 8, 2001 [X] [2] Original Deposit [ ] [2] New Deposit [ ] [2] Repropagation of Original Deposit

III. (a) VIABILITY STATEMENT

Deposit was found: [X] Viable [ ] Nonviable on May 10, 2001 (Date)

International Depositary Authority's preparation was found viable on May 18, 2001 (Date)[3]

III. (b) DEPOSITOR'S EQUIVALENCY DECLARATION

Depositor determined the International Depositary Authority's preparation was

[X] [2] Equivalent [ ] [2] Not equivalent to deposit on 7/23/01 (Date)

Signature of Depositor Steve Hill

IV. CONDITIONS UNDER WHICH THE VIABILITY TEST WAS PERFORMED (Depositors/Depositary)[4]

V. INTERNATIONAL DEPOSITARY AUTHORITY

Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street Peoria, Illinois 61604 U.S.A.	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s): J Swyer Date: 6-26-01

[1] Indicate the date of the original deposit or when a new deposit has been made.
[2] Mark with a cross the applicable box.
[3] In the cases referred to in Rule 10.2(a)(ii) and (iii), refer to the most recent viability test.
[4] Fill in if the information has been requested.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO
Syngenta
3054 Cornwallis Road
Research Triangle Park, NC 27709

VIABILITY STATEMENT

issued pursuant to Rule 10.2 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF THE PARTY TO WHOM THE VIABILITY STATEMENT IS ISSUED

I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM
Name: Syngenta Address: 3054 Cornwallis Road Research Triangle Park, NC 27709	Depositor's taxonomic designation and accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: *Pseudomonas putida* NRRL B-30479 Date of: May 8, 2001 [X] [2] Original Deposit [ ] [2] New Deposit [ ] [2] Repropagation of Original Deposit

III. (a) VIABILITY STATEMENT

Deposit was found: [X] Viable [ ] Nonviable on May 10, 2001 (Date)

International Depositary Authority's preparation was found viable on May 16 2001(Date)[3]

III. (b) DEPOSITOR'S EQUIVALENCY DECLARATION

Depositor determined the International Depositary Authority's preparation was

[X] [2] Equivalent [ ] [2] Not equivalent to deposit on 7/23/01 (Date)

Signature of Depositor Steve Hill

IV. CONDITIONS UNDER WHICH THE VIABILITY TEST WAS PERFORMED (Depositors/Depositary)[4]

V. INTERNATIONAL DEPOSITARY AUTHORITY

Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street Peoria, Illinois 61604 U.S.A.	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s): Date: 6-26-01

[1] Indicate the date of the original deposit or when a new deposit has been made.
[2] Mark with a cross the applicable box.
[3] In the cases referred to in Rule 10.2(a)(ii) and (iii), refer to the most recent viability test.
[4] Fill in if the information has been requested.